npdc

NAVY PERSONNEL RESEARCH AND DEVELOPMENT CENTER, SAN DIEGO, CALIFORNIA 92152

NPRDC TR-74-29 FEBRUARY 1974

Illinois University.

DEVELOPMENT AND IMPLEMENTATION OF THE COMPUTER ASSISTED INSTRUCTION STUDY MANAGEMENT SYSTEM (CAISMS)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: TR 74-29
2. GOVT ACCESSION NO.:
3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): DEVELOPMENT AND IMPLEMENTATION OF THE COMPUTER ASSISTED INSTRUCTION STUDY MANAGEMENT SYSTEM (CAISMS)
5. TYPE OF REPORT & PERIOD COVERED: Technical Report, 1 January 73 - 1 January 74
6. PERFORMING ORG. REPORT NUMBER:

7. AUTHOR(s):
Stephen M. Allessi
Richard C. Anderson
Thomas H. Anderson
W. Barry Biddle
Bruce R. Dalgaard
Donald W. Paden
H. Richard Smock
John R. Surber
Edward J. Wietecha

8. CONTRACT OR GRANT NUMBER(s): N61339-73-C-0078

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
University of Illinois
Urbana-Champaign, Illinois 61801

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:

11. CONTROLLING OFFICE NAME AND ADDRESS:
Navy Personnel Research and Development Center
San Diego, California 92152

12. REPORT DATE: February 1974
13. NUMBER OF PAGES: 126
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):
15. SECURITY CLASS. (of this report): UNCLASSIFIED
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report):

Approved for Public Release; Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):

18. SUPPLEMENTARY NOTES:

Prepared in cooperation with Navy Personnel Research and Development Center.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number):

computer assisted instruction
computer management of instruction
instructional technology
prose learning
computer based testing

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper reports the work accomplished during the initial year of a contract between the University of Illinois and Navy Personnel Research and Development with Advanced Research Projects Agency support. The purpose of the project was to design, tryout, and evaluate a system for maintaining attentive study of instructional materials. A CAI system was used for this purpose but, in contrast to most CAI efforts, existing materials were used and students spent minimal time in on-line contact with the computer. The

report includes a manual of procedures for preparing test items which maintain attentive study, evaluation of the system, cost projections for use of the system, and a suggested extention of the system.

NPRDC TR 74-29 February 1974

DEVELOPMENT AND IMPLEMENTATION OF THE COMPUTER ASSISTED INSTRUCTION STUDY MANAGEMENT SYSTEM (CAISMS)

Stephen M. Allessi
Richard C. Anderson
Thomas H. Anderson
W. Barry Biddle
Bruce R. Dalgaard
Donald W. Paden
H. Richard Smock
John R. Surber
Edward J. Wietecha

University of Illinois at Urbana-Champaign

This research was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by Navy Personnel Research and Development Center under Contract NO. N61339-73-C-0078.

Reviewed by
William E. Montague

Approved by
James J. Regan
Technical Director

Prepared for

Navy Personnel Research and Development Center
San Diego, California 92152

FOREWORD

This research was performed under Exploratory Development Task Area PF55.522.002 (Methodology for Developing/Evaluating Navy Training Programs) Work Unit Number PF55.522.002.01.60 (Advanced Computer Based Research). This Work Unit is jointly funded by The Advanced Research Projects Agency (Account Symbol 9740400.1311). The motivating requirements for this project are for "... improvements in training methodologies, measurement techniques, management and administration, including decision criteria required for their rapid implementation" contained in General Operational Requirement 43, Rev. 10/71.

Dr. William E. Montague was the technical contract monitor for the Navy Personnel Research and Development Center on this research effort. The basic rationale for the research was developed in a proposal submitted by Dr. R. C. Anderson during the Fall of 1972.

F. L. Nelson
Commanding Officer

SUMMARY

Problem

This report summarizes the developmental phase of a project designed to implement a course of study on a computer-assisted instruction (CAI) system maximizing instructional effectiveness while minimizing both time required to prepare materials and student contact time on the CAI system. Since contact time was to be minimized, the immediate problem confronted was to maintain control over the students' learning. Students often fail to learn from text materials because they don't study carefully enough. Therefore, this project attempted to design, tryout, and evaluate a practical CAI system which would maintain attentive study of instructional materials when the student was off-line. Existing materials were employed to minimize effort in materials preparation. Prior research had indicated that: (1) a series of processing stages are required for learning from text, (2) these processing stages deteriorate under self-study conditions, (3) procedures which maintain full processing dramatically facilitate learning, and (4) the technique of frequent, intermittent questioning can be used to maintain control of the learning when the student is studying off-line.

Approach

The CAI study management system (CAISMS) was designed to question the student intermittently about what he is reading in order to maintain deep processing. The system was implemented on the PLATO CAI system at the University of Illinois. Terminals were installed in the library to be near an environment in which students could study. The instructional sequence involved the student signing in on the terminal, receiving an assignment, studying the material in the nearby workspace, signing in again and receiving a quiz covering the reading. If successful, the next assignment was given and the sequence was repeated. Procedures were developed to generate questions from the text materials, and a computer program was written to select and display questions about the assignment, evaluate and record responses, and keep track of student progress. Performance was evaluated by scores on the quizzes and by off-line examinations. Daily records were kept on student interaction with the terminals in order to monitor studying behavior. Student attitudes toward the course were assessed by means of a questionnaire. A comparative study of the effect of CAISMS was not undertaken during this developmental phase. An experimental comparison is currently (Spring 1974) underway.

Findings and Conclusions

CAISMS proved to be simple to install, relatively inexpensive to operate, convenient to use, and is compatible with a variety of educational and training environments. Achievement, procedural and attitudinal data

indicate that the study management technique is feasible to administer and potentially effective in producing achievement gains over traditional ongoing instruction. Most students had favorable to very favorable opinions of the system. However, they were not favorably disposed to the idea of having to study near a terminal in the library.

TABLE OF CONTENTS

SECTION		Page
I.	THE RATIONALE FOR CAISMS	1
II.	THE PREPARATION OF TEST ITEMS TO MAINTAIN ATTENTIVE STUDY BEHAVIOR	23
III.	THE DEVELOPMENT OF A COMPUTER PROGRAM TO DELIVER CAISMS.	50
IV.	EVALUATION OF CAISMS	63
V.	TOWARDS A STUDENT RECORD KEEPING SYSTEM	72
VI.	SOME REFLECTIONS ON THE USE OF CAISMS	78
VII.	AN ECONOMIST'S PERSPECTIVE ON CAISMS	82
VIII.	COST PROJECTIONS ON THE DEVELOPMENT AND USE OF CAISMS	88
IX.	CAISMS II: A PROPOSED INNOVATIVE COURSE MANAGEMENT SYSTEM	94

TABLES

Page

II.1. Number of Subsections of Text, Item Format, and the Number and Type of Items Constructed During First Development Effort 34
II.2. Number of Subsections of Text, Item Format, and the Number and Type of Items Constructed During Second Development Effort 46
IV.1. Some Descriptive Statistics of the Sample of Economics Students 64
IV.2. A Listing of Eleven Predictor Variables and Regression Coefficients Used in a Significant Canonical Regression Equation with Three Exam Scores as Criteria 69

FIGURES

II.1. The process for constructing questions 32
II.2. Subsection breakdown for Chapter 1 of the text, The American Nation, indicating the subsection title, code number and page references 40
II.3. The record keeping procedure used for the development of test items for Chapter 1 of The American Nation . . . 42
II.4. Identifying code number, subsection title, page reference numbers and summary statement 44
III.1. A flow diagram and description of the computer based instructional process used in CAISMS 54
IV.1. A cumulative study progress record of fifteen economics students 66
V.1. A hard copy report of student performance on a cluster of four items 73
V.2. A hard copy report of a student's progress in CAISMS . . . 74
VIII.1. Cost per student for three credit hours of instruction . . 92

I. THE RATIONALE FOR CAISMS

Richard C. Anderson

The purpose of the research and development project described in this report has been to design, tryout, and evaluate a practical computer assisted instruction (CAI) system for maintaining attentive study of instructional materials. Unlike most CAI projects, this one employs existing materials and the student or trainee spends only a fraction of his time on-line.

The idea is very simple: frequently students fail to learn from books because they don't study them carefully enough. The study management system is designed to intermittently question the student about what he is reading so as to maintain deep processing. The student signs in at the computer terminal and receives a brief assignment. Upon completing the assignment in a nearby work space, the student again signs in. This time he receives a short quiz over the assignment just finished. Then the cycle starts again with the next assignment.

The CAI study management system (CAISMS) has proved to be simple to install, relatively inexpensive to operate, convenient to use, and we believe it will prove compatible with a variety of instructional modes in a variety of educational and training environments. Relative to unmonitored study, increments in achievement have been observed. Relative to fully on-line CAI, large savings in development costs and somewhat smaller savings in operating costs seem probable.

The rationale for the claim that a CAI study management system will be effective contains four parts. (1) A series of processing stages is

required for learning from printed verbal-symbolic matter. (2) These processing stages are likely to deteriorate under the conditions which prevail during self-study. (3) Procedures which maintain full processing dramatically facilitate learning. (4) The technique of intermittent questioning is a high probability candidate for practical management of processing activities. Each of these contentions will be considered in turn.

Processing Stages

There appear to be at least three processes, or encoding stages, necessary for a person to learn from printed verbal material. First is a perceptual encoding stage. Since the relevant perceptual features of text are orthographic, this process can be said to give rise to an orthographic code. It is usually assumed that this code has a very short memorial life. The next processing stage involves acoustical or phonological features; the printed symbols are translated into implicit (or explicit) speech. The phonological code has been identified as the chief form of representation in short term, or primary, memory. A third stage, and a vital one I shall argue, entails semantic encoding--in common language, bringing to mind a meaningful representation for the words. The presumption is that the semantic code is the principal form of long term storage. For more extensive discussions of stage processing views of learning see Neisser (1967), Anderson (1970), Montague (1972), and Craik and Lockhart (1972).

The important practical implication of the encoding stage theory is that a person usually will get the information in a communication into long-term storage if he semantically encodes it, and usually will not get the information into long-term storage if the communication has received a

shallower level of processing. Two studies will be cited in support of this important proposition:

Bobrow (1970) presented pairs of nouns twice, each time embedded in a sentence (e.g., The animal's bark scared the big league pitcher). On the test the first noun from each sentence (e.g., bark) was presented as a retrieval cue; the subject had to give the second noun (e.g., pitcher) as a response. Changes in the second sentence did not affect recall as long as the referents of the nouns remained the same (e.g., The dog's bark frightened the baseball pitcher). However, when the sense of the nouns changed (e.g., The medicinal bark filled the porcelain pitcher), recall was substantially lower. If the people were storing phonological codes, context-induced changes in the meanings of the nouns would not have affected performance. Hence, the results imply that meanings rather than sounds were stored. Anderson (1971) completed two experiments in which college undergraduates attempted to recall the subject noun of each of a series of once exposed sentences. As retrieval cues they were given the verbatim predicate or its paraphrase. The paraphrased predicates contained no substantive words in common with the originals. For instance, one of the sentences was The traveler appreciated the gift. The related test items were The ________ appreciated the gift and The ________ was grateful for the present. The conditional probability of a correct response to a paraphrase item, given a correct response to the verbatim item based on the same sentence, was nearly .90. This is strong evidence that learning from a sentence usually entails semantic encoding, because a paraphrase and its base are related in terms of meaning but unrelated

with respect to the shape or the sound of the words. A person could not answer a paraphrase question unless he had comprehended the sentence.

While it appears that long term storage usually entails semantic encoding, there is experimental evidence that verbal material can be stored for the long term in an acoustical or phonological code (e.g., Bruce & Browley, 1970). Common experience leads to the same conclusion. We all have learned a poem or a song in a foreign language, which we can rattle off in perfect form, though we don't have the slightest idea of the meaning of the words we are saying. It would appear, then, that at least some of the time under some conditions people put the surface form of communications into long term storage.

The study of semantics is in its infancy. Hence, it is impossible to be very precise about what a person has in his head when he can be said to have the meaning. This fact is not a serious obstacle to the present project, even though the notion of semantic code, and rough equivalents from ordinary language such as comprehension and understanding, is a concept upon which we shall rely heavily. As the author (Anderson, 1972, p. 148) has said recently,

> It is not necessary to know the precise form and the mental organization of meanings in order to recognize the occurrence and nonoccurrence of comprehension. The trick will be to devise techniques for constructing questions that can be answered if a person has semantically encoded a communication but not answered if it has been encoded only perceptually or phonologically.

In summary, a series of processing activities seems to be required in order for a person to learn from printed instructional materials. Ernst Rothkopf (1966, 1968, 1970) has coined the neat term mathemagenic to refer

to these activities. According to Rothkopf (1970) the term is derived from the Greek roots mathemain--that which is learned--and gignesthai--to be born. Mathemagenic behaviors are the behaviors which give birth to learning.

Deterioration of Processing

Our general thesis is that the activities the student or trainee engages in when confronted with instructional tasks are of crucial importance in determining what he will learn. We reject the long-dominant alternative view that the student is a passive receptacle whose learning and performance are directly determined by input variables.

Unfortunately, it seems probable that students, even successful ones (Danskin & Burnett, 1952), are quite inefficient at studying. It appears that essential processing activities attenuate sharply when the material is difficult or lengthy or the student is tired, bored, distracted, or under pressure to work quickly. These conclusions are based primarily on questionnaire investigations of study habits. A review of much of this research along with a sizeable bibliography can be found in Effective Study, the book by Francis Robinson, the well-known how-to-study authority.

Research involving systematic observation or experimental investigation of processing activities that has utilized real instructional materials under life like conditions has been rare. The research that has been done confirms that mathemagenic activities are prone to disruption.

Hoffman (1946) completed a suggestive study in which college undergraduates read a history textbook for four consecutive hours. Eye movements were monitored by recording changes in electrical potential through

electrodes attached near the reader's eyes. The "electroculograms" showed that the pattern of eye movements associated with reading steadily deteriorated over the four-hour period. For instance, half-hour by half-hour there was a regular increase in the frequency of eye blinks and the variability of eye movements, and a regular decrease in the number of lines read per minute. Hoffman's graphs resemble extinction curves.

Anderson (1970) has reviewed evidence which shows that the consistent failures to find benefits from prompting techniques, the requirement to make active overt responses, and immediate feedback in programed lessons can all plausibly be attributed to "short circuiting" of attention and processing.

We find appealing the somewhat speculative notion that processing deteriorates from "inside to out" (see Rothkopf, 1968). By this we mean that probably the first process to drop out is thinking that goes beyond the text. Then semantic encoding may stop. For the reader this is the deceptive twilight zone between reading and not reading. His eyes are tracking lines of print. He is translating the written words into speech. But there is little cognitive involvement.

Everyone has had the experience of "reading" several pages, only suddenly to realize that he has no more than a vague impression of what has passed before his eyes. About persons who have reached the stage when semantic encoding has decayed while lesser processes continue, Robinson (1961, pp. 13-14) has written:

> One almost gets an impression of dutiful line following so that the next day they can truthfully say 'I don't remember, but, honest, I read every word.' ... As they read along they can continually murmur 'mmhm,' 'uhfmm' as they see each

> idea, much as a mirror passing over the book might clearly reflect what was printed. On finishing, they push the book aside with a sigh. To an impolite inquiry as to what ideas were discussed, the typical reader has a nebulous feeling that there was much he had understood but it now is jumbled.

Next, the phonological and orthographic encoding processes may deteriorate. The reader may begin to "skim." The frequency of blinks and regressive eye movements increase. The reader's mind begins to wander. He becomes increasingly vulnerable to distractions. He may begin to skip huge chunks of text. He may go to sleep. He may find he "needs" a cigarette or a cup of coffee. He may "discover" that an entirely different task "requires" his attention.

Parenthetically, we are not proposing a remedy for really gross failures of exposure to instructional stimuli. Other techniques, such as contingency management procedures, have shown great promise for keeping the student or trainee "on task" (see, for instance, Becker, Madsen, Arnold, & Thomas, 1967). The CAI study management system is aimed at the person who is at least superficially on task, and who is giving at least minimal compliance to the demands of the instructor or instructional agent. One potentially valuable side benefit of the study management system will be the day by day identification of persons who fail to show up for instruction, or do so erratically. Up to date cumulative records of the rate and extent of each individual's study behavior will be a by product of the system, as will be detailed later.

In summary, it is conjectured that the activities required to give birth to learning deteriorate in the following order:

1. Failures of higher-level processing that goes beyond the text.

2. Failures of meaningful processing.

3. Failures of speech processing and perceptual tracking.

4. Gross failures of exposure to instructional stimuli.

It is assumed, and the available evidence seems to indicate, that the extinction-like decay of mathemagenic activities is inevitable even for the finest student during his better moments, but that the deterioration becomes increasingly swift and certain when the material is difficult or lengthy, or the student is fatigued, unmotivated, of low ability, or under time pressure.

Effects of Procedures that Induce Meaningful Processing

If semantic encoding is essential for learning but such encoding is quite likely to "drop out" when a person studies, it follows that procedures which induce meaningful processing will sharply facilitate learning. Recent research confirms this expectation.

Bobrow and Bower (1969) instructed subjects in one group to compose a sensible continuing sentence for each sentence they say. This group recalled over twice as many words as a control group required to read each sentence aloud three times. To compose a sensible continuing sentence required comprehension, that is, semantic encoding, of the presented sentence whereas reading it aloud again and again did not.

Anderson, Goldberg, and Hidde (1971) constructed sentences in which the last word of each was determined by the rest of the sentence. Subjects who supplied the missing last word in each sentence as they read aloud learned more than subjects who read aloud whole sentences. One of the incomplete statements was, Circles have neither a beginning nor an ____

To complete the sentence with the word "end" required a person to bring to mind a meaningful representation for the rest of the sentence. "End" is semantically rather than phonologically related to the rest of the sentence; hence, simply rendering the printed words into speech would not allow a person to fill the blank correctly.

Anderson and Kulhavy (1972b) prepared one-sentence definitions of a set of rare words in the back of Thorndike and Lorge (1944) which several judges said they did not know. Also prepared were two multiple-choice test items for each word. In each item the rare word was listed; the person had to select an example of the concept named by the word from among four alternatives. The description of the examples did not contain substantive words in common with definitions. Following one paced exposure of the definitions, college undergraduates instructed to use each defined word in a sensible sentence as the definitions appeared answered one-and-a-half times as many questions correctly as those who read each definition aloud three times. Again the explanation is that a person can "read," that is, translate printed symbols into speech sounds, without bringing to mind the meaning of the words he is speaking. But to compose a sensible sentence requires comprehension.

Preliminary instructions can affect the character of subsequent processing and thereby influence learning. Instructions to "read carefully" usually have a slight facilitative effect on learning from prose (Rothkopf, 1966; Brunning, 1968). Somewhat stronger effects accrue from telling people to read brief passages for gist or substance, even when recall is scored verbatim (Cofer, 1941; King & Russell, 1966). Strongest of all are

the effects that can be obtained with imagery instructions. Subjects instructed to form a vivid mental image of the situation described in each of a series of unrelated sentences recall about three times as much as subjects who read each sentence aloud several times (Anderson & Hidde, 1971; Anderson, 1971). Obviously the person who takes seriously the instruction to form an image will be engaging in semantic encoding.

In summary, the research demonstrates that diverse procedures, which appear to have in common only that they induce meaningful processing, strongly facilitate learning. None of the techniques reviewed thus far is a good candidate for maintaining deep processing of a large volume of actual text material, however. An exhortation to learn main ideas, for instance, or to form images is likely to lose its effect after a while. In fact, Anderson and Kulhavy (1972a) were unable to show any improvement from imagery instructions in learning a 2,000 word prose passage.

Adjunct Questions

That asking students questions now and then will improve learning is an embarrassingly unoriginal idea! However, our claim is not that the technique is original but rather that it is effective. Those who take the trouble to search the literature will find the positive effects of "test-like events" on learning and retention probably the best documented proposition in the field. Studies reviewed by Gates (1917) show the effects were already known just after the turn of the century. Gates' own research showed substantial benefits from "active recitation" on the learning and remembering of both serial lists of nonsense syllables and brief prose passages. Jones (1923) had subjects read three brief prose

selections and immediately thereafter complete one of two tests covering the selections. A day later all subjects took both tests. Scores on the repeated test were twice as high as scores on the test taken for the first time. Numerous investigators since then have corroborated the facilitative effects of questions during or shortly after instruction. Of course, the recent wave of interest in questioning has been stimulated by the research of Rothkopf.

Both direct and indirect effects of questioning have been demonstrated. By "direct effect" we mean the increment in performance which is observed when a question asked during or shortly after instruction is repeated on a later test. The direct effect is usually large. Increments in the range of one-and-a-half to two times the mean of a reading-only control group are not uncommon. Rothkopf (1966) can be credited with proving that inserted questions also have indirect effects. On new posttest questions, shown experimentally to be unrelated to the inserted questions, groups that receive the inserted questions score significantly higher than reading-only control groups. Since 1966 at least a half dozen experiments by several different investigators have confirmed that adjunct questions have small but consistent indirect effects.

There is converging evidence that much of the indirect effect of adjunct questions is due to maintenance and shaping of processing activities. For instance, Rothkopf and Bisbicos (1967) presented two questions after each three-page section of a thirty-six page passage on marine biology. Some subjects saw only questions requiring either a measured quantity (e.g., a distance or a date) or a proper name for an answer.

Some saw only questions which could be answered with a common English word or a technical term (e.g., bathyscaphe, phototropic). On the criterion test, subjects did relatively better than appropriate controls on new questions from the same category as they had received during learning and relatively worse on questions from other categories. The straightforward interpretation is that the questions selectively focused attention on certain classes of information.

Recently McGaw and Grotelueschen (1972) have discovered that adjunct questions have a "backward" review influence as well as the "forward" shaping influence discussed above. They found that answering a question facilitated recall of the material in the vicinity of the sentence upon which the adjunct question was based. The design and procedures ruled out the possibility that this result was due to the direct or transfer effects of the question, leaving as the only sensible explanation that subjects covertly reviewed the surrounding material while retrieving and verifying the answer to the adjunct question itself. Rothkopf and Billington (unpublished) have confirmed these results and interpreted them in the same fashion.

Briefly considered next are variables and procedures which influence the magnitude of the influence of adjunct questions.

Position of questions. The direct effect of questions does not appear to depend critically upon the position of the questions relative to the text. However, the indirect effects are obtained only when the questions appear after the segment of text upon which they are based. The data and reasoning upon which these conclusions are based have been

detailed by Frase (1970). It appears that the indirect review effect is obtained only on material appearing shortly before the adjunct questions (McGaw & Grotelueschen, 1972).

Amount of material before questions. The research done to date suggests that the best policy is a few questions after short assignments (Webb & Schwartz, 1959; Frase, 1968). However, questions can have strong effects even when they follow fairly lengthy assignments. Anderson and Myrow (1971), using 2,200-word passages describing fictitious primitive tribes, found that groups given an immediate test averaged 71% when the test was repeated a week later whereas groups that had not received the immediate test averaged only 50%. Roderick and Anderson (1968) got comparable results in a study involving a 3,400-word self-instructional program on classical conditioning and a 1,800-word textbook-like summary of the same material.

Interval between study and question-answering. How long an interval should there be after a student has studied instructional materials before he is tested? The answer seems to be that the sooner the questions are asked the better. Spitzer (1939) had nine groups of sixth graders, totalling 3,600 students, read a 577-word passage describing an agriculture experimental station. At various times thereafter the groups received a 25-item multiple choice test. Feedback was never given. Groups that took the test for the first time immediately after reading the passage showed virtually no forgetting for as long as seven weeks. Groups tested for the first time 24 hours after reading had already forgotten 50% of the information learned from the text, though the test taken at this point

did sustain performance on later tests. In general, the effect of the test was to "preserve" what the student could remember at that point but it was not effective in helping him "recover" what he had already forgotten.

Difficulty of questions. We know of two studies which have examined the effects of test difficulty on subsequent student performance. Sax and Reade (1964) gave half the students in an educational measurement course two unit exams with a difficulty index of about .40. The remainder got unit exams with a difficulty index of .70. Those who had received the more difficult exams did significantly better on a final examination of intermediate difficulty. Marso (1969) in a rather similar study got opposite results; those who received easy unit exams scored higher on the final. Marso completed a second study in which he found that people given easy unit exams reported subsequently spending more hours studying than people who received difficult unit exams. For some reason, Marso did not report the results on the final examination in the second experiment.

We find most creditable the lines of reasoning which point to the conclusion that relatively easy questions facilitate subsequent performance, but obviously this is an issue which will have to be investigated further.

Type of question. We come now to a crucial issue, the type of question. There is no way to tell for sure what types of questions have been used in the bulk of the investigations completed to date. The reason is simply that the research reports have not described the questions very thoroughly. Most investigators provide only such rudimentary information as the number of questions and the response mode (e.g., multiple choice,

short answer). In the cases in which the adjunct questions and criterion tests have been explicitly described, the questions usually have consisted of statements lifted in essentially verbatim form from the text. A word or phrase--usually a proper name, technical term, or measured quantity--has been deleted from each statement, to be supplied by the subject. Studies employing this kind of question have yielded positive results, but consider these two facts: (1) Few instructional goals are well-served by the capacity for verbatim recall. Important objectives most often entail using skills, identifying new instances of concepts, applying principles, and so on. (2) If the encoding stage theory outlined earlier is correct, questions which require deep semantic processing will facilitate learning to a greater extent than the verbatim recall questions that have apparently predominated in previous research.

Watts and Anderson (1971) had 300 high school seniors answer a question after reading each of five 450-word passages explaining a psychological principle. The question involved identifying an example of the principle or the name of the psychologist associated with the principle. Subjects who received questions that required them to apply the principles to new examples performed significantly better on the posttest than all other subjects, including subjects who received otherwise identical questions that repeated in verbatim form examples described in the text. The groups who answered application-to-new-example questions while reading were at no disadvantage on posttest repeated-example and name questions, while at the same time demonstrating a striking superiority on both repeated and alternate application questions. It is pretty clear

that a person could answer a repeated-example question by matching the surface phonological or orthographic form of the correct answer with a surface code for the relevant piece of text. A person had to <u>understand</u> a principle in order to apply it to a new example. The implication is that to reach any but trivial instructional objectives one must ask questions which assuredly require comprehension to answer.

<u>An Augmented Theory of Questioning</u>

As will be explained in detail later in this report, our initial assumption was that the best questions to ask students were ones that would require comprehension, namely, paraphrase and application questions. Research completed during the course of the project has raised doubts about this assumption. Anderson, Surber, Biddle, Zych, and Lieberman (1973) found to their great surprise that people who received a verbatim quiz did slightly better on a delayed test than people who received a paraphrase quiz. This fact caused us to revise our theory. It is now argued that the process by which a quiz question enhances delayed retention involves two stages. First, the question must permit retrieval of information from short term memory. Second, the question must instigate meaningful processing of the information so as to transfer it into long-term, semantic memory.

This theory can be expressed in the following equation,

$$P(D) = k + (1-k)\, rt,$$

which says that the probability of a correct response on the delayed test, P(D), equals the proportion of items of information the person already knows, $\underline{k}$, either information he knows beforehand or learns from reading

the passage, plus an increment due to taking the quiz. The increment consists of that information not already known which the questions cause the person to retrieve, r, from short term, phonological memory and transfer, t, into long term, semantic memory.

As expected on the basis of the augmented theory two experiments involving 662 high school students suggested that verbatim items were very much better at evoking retrieval of information whereas paraphrase items were better at instigating transfer. What we had overlooked originally was that information in short term memory is phonologically coded, making it accessible to a verbatim question but relatively inaccessible to a paraphrased one. The second experiment confirmed a nonobvious prediction from the theory: a verbatim quiz followed by a paraphrase quiz optimumly facilitates delayed retention. We call this the "piggyback" treatment.

Currently we are testing the practical value of the piggyback treatment. On the basis of data gathered during the first semester of operation a number of difficult economics questions were identified. After screening out items that contained ambiguities or other difficulties affecting validity, a verbatim question was written to parallel each remaining difficult item. The items have been divided into two blocks. Half of the students currently enrolled in the CAISMS sections of Economics 108 receive a piggyback verbatim item just prior to each target item in the first block but no such treatment before the items in the other block. The assignment is reused for the remaining students.

Summary

In summary, 75 years of research have consistently shown that requiring students to answer questions facilitates learning. Questions appear to (1) maintain and shape processing of subsequent material and to (2) encourage additional and sometimes deeper processing of previously encountered material, provided this material still has at least a fragmentary representation in memory. While a few issues remain in doubt, the conditions required in order for questioning to improve learning appear to be sufficiently well understood to make practical apppication a good bet at this time.

References

Anderson, R. C. Control of student mediating processes during verbal learning and instruction. Review of Educational Research, 1970, 40, 349-369.

Anderson, R. C. Encoding processes in the storage and retrieval of sentences. Journal of Experimental Psychology, 1971, 91, 338-340.

Anderson, R. C. How to construct achievement tests to assess comprehension. Review of Educational Research, 1972, 42, 145-170.

Anderson, R. C., Goldberg, S. R., & Hidde, J. L. Meaningful processing of sentences. Journal of Educational Psychology, 1971, 62, 395-399.

Anderson, R. C., & Hidde, J. L. Imagery and sentence learning. Journal of Educational Psychology, 1971, 62, 526-530.

Anderson, R. C., & Kulhavy, R. W. Imagery and prose learning. Journal of Educational Psychology, 1972, 63, 242-243. (a)

Anderson, R. C., & Kulhavy, R. W. Learning concepts from definitions. American Educational Research Journal, 1972, 9, 385-390. (b)

Anderson, R. C., & Myrow, D. L. Retroactive inhibition of meaningful discourse. Journal of Educational Psychology Monograph, 1971, 62, 81-94.

Anderson, R. C., Surber, J. R., Biddle, W. B., Zych, P. M., & Lieberman, C. E. Retention of text information as a function of the nature, timing, and number of quizzes. Paper read at the annual convention of the American Educational Research Association, Chicago, 1974.

Becker, W. C., Madsen, C. H., Arnold, C. R., & Thomas, D. R. The contingent use of teacher attention and praise in reducing classroom behavior problems. Journal of Special Education, 1967, 1, 287-307.

Bobrow, S. A. Memory for words in sentences. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 363-372.

Bobrow, S. A., & Bower, G. H. Comprehension and recall of sentences. Journal of Experimental Psychology, 1969, 80, 455-461.

Bruce, D., & Crowley, J. J. Acoustic similarity effects on retrieval from secondary memory. Quarterly Journal of Experimental Psychology, 1970, 9, 190-196.

Brunning, R. H. Effects of review and test-like events within the learning of prose material. Journal of Educational Psychology, 1968, 59, 16-19.

Cofer, C. N. A comparison of logical and verbatim learning of prose passages of different lengths. American Journal of Psychology, 1941, 54, 1-20.

Craik, F. I. M., & Lockhart, R. S. Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 671-684.

Danskin, D. G., & Burnett, C. W. Study techniques of those superior students. Personnel and Guidance Journal, 1952, 31, 181-186.

Frase, L. T. Effect of question location, pacing, and mode upon retention of prose material. Journal of Educational Psychology, 1968, 59, 244-249.

Frase, L. T. Boundry conditions for mathemagenic behaviors. Review of Educational Research, 1970, 40, 337-348.

Gates, A. I. Recitation as a factor in memorizing. Archives of Psychology, 1917, 7 (Whole No. 40).

Hoffman, A. C. Eye movements during prolonged reading. Journal of Experimental Psychology, 1946, 36, 95-118.

Jones, H. E. The effect of examination on permanence of learning. Archives of Psychology, 1923-24, 10 (68), 1-70.

King, D. J., & Russell, G. W. A comparison of rote and meaningful learning of connected, meaningful material. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 478-483.

Marso, R. N. The influence of test difficulty upon student achievement. American Educational Research Journal, 1969, 6, 621-632.

McGaw, B., & Grotelueschen, A. The direction of the effect of questions in prose material. Journal of Educational Psychology, 1972, 63, 580-588.

Montague, W. E. Elaborative strategies in verbal learning and memory. In G. H. Bower (Ed.), The psychology of learning and motivation. Vol. 6. New York: Academic Press, 1972.

Neisser, U. Cognitive psychology. New York: Appleton-Century-Crofts, 1967.

Robinson, F. Effective study. New York: Harper and Row, 1961.

Roderick, M. C., & Anderson, R. C. A programed introduction to psychology versus a text-book style summary of the same lesson. Journal of Educational Psychology, 1968, 59, 381-387.

Rothkopf, E. Z. Learning from written materials: An exploration of the control of inspection behavior by test-like events. American Educational Research Journal, 1966, 3, 241-249.

Rothkopf, E. Z. Two scientific approaches to the management of instruction. In R. M. Gagné and W. J. Gephart (Eds.), Learning research and school subjects. Itasca, Ill.: Peacock, 1968. Pp. 107-132.

Rothkopf, E. Z. The concept of mathemagenic activities. Review of Educational Research, 1970, 40, 325-336.

Rothkopf, E. Z., & Billington, M. J. Indirect review effects produced by adjunct questions: A replication study. Unpublished manuscript.

Rothkopf, E. Z., & Bisbicos, E. Selective facilitative effects of interspersed questions on learning from written materials. Journal of Educational Psychology, 1967, 58, 56-61.

Sax, G., & Reade, M. Achievement as a function of test difficulty level. American Educational Research Journal, 1964, 1, 22-25.

Spitzer, H. F. Studies in retention. Journal of Educational Psychology, 1939, 30, 641-656.

Thorndike, E. L., & Lorge, I. The teacher's word book of 30,000 words. New York: Columbia University, Teacher's College Press, 1944.

Watts, G. H., & Anderson, R. C. Effects of three types of inserted questions on learning from prose. Journal of Educational Psychology, 1971, 62, 387-394.

Webb, W. B., & Schwartz, M. Measurement characteristics of recall in relation to the presentation of increasing large amounts of material. Journal of Educational Psychology, 1959, 50, 63-65.

II. THE PREPARATION OF TEST ITEMS TO MAINTAIN ATTENTIVE STUDY BEHAVIOR

Edward J. Wietecha and Richard C. Anderson

This section of the report discusses the procedures used in constructing sets of test items for assessing student comprehension of prose text material. Theoretical discussion and empirical data on question asking can be found in Bormuth (1970), Anderson (1972) and Section I of this report.

The Development of Summary Statements

An underlying assumption of the procedure discussed in this section of the report is that the prose text material contains the information necessary for the student to master a concept or to gain a basic understanding of a topic. Furthermore, it is assumed that the text conveys that information to the student in a manner which the student is able to read and understand. Once these assumptions have either been met, or accepted as being met, the important ideas and/or concepts are abstracted from the text through the development of summary statements.

A unit of analysis is decided upon by the item writing team. In some situations the unit of analysis might be a paragraph of text material, a page of material or it might be subsections of a chapter of the text. For each unit of analysis, at least one statement is written, in the verbatim language of the text, which summarizes a main idea. A subject matter expert is necessary for this phase of development. Certain sections of the text such as introductory paragraphs or sections of summary material at the end of a chapter will usually not have a summary statement

written for them. In some cases, the subject matter expert will decide that a section or paragraph does not contain an important idea and will not construct the summary statement. Conversely some sections of the text may be relatively "meaty" and several summary statements will have to be constructed.

As stated above, the summary statements should be taken from the text in verbatim form. This may mean lifting an entire sentence from the text, or it may mean piecing together key phrases from several sentences. The substantive words (nouns, verbs, modifiers) should literally be taken from the text, but the syntax and tense may be modified if necessary. To avoid ambiguity in the summary statement, pronouns should be replaced with the appropriate noun or phrase.

Under each summary statement the examples used to illustrate the concept or idea should be noted. Only the examples in the specific unit of analysis need be listed.

Preparation of Test Items

A test item should be constructed, or selected from an existing pool of items, for each summary statement. The summary statement and identification code should be written at the top of a sheet of paper and the corresponding item written onto the page. Each summary statement and corresponding test item should be written on a separate page.

In order to discount rote learning of the surface form of the text, the language of the test items and the text should be as different as possible. That is, synonymous words (or phrases) should be substituted for the language of the original. This rule applies to all test items,

including ones that involve applying principles to new examples. The language of such items should not overlap excessively with either the general statement of the principle or with the instructional examples. Ideally, the test item and the text will have no substantive words (nouns, verbs, modifiers) in common.

One exception to the paraphrase rule is the sort of item in which the student is expected to select the definition of a concept or principle given the name of the concept or principle. Obviously it would be inappropriate to modify the term being defined.

The 'new case' rule. If the statement expresses a concept, explains a principle, or describes a procedure, the preferred strategy is to prepare a test item which requires the student to apply this knowledge to a new case.

A "new case" is one different from any used as an illustration. Cases vary in similarity to instructional examples. The appropriate test case is the most remote one the student could reasonably be expected to handle on the basis of the text.

When constructing an item that calls for application of knowledge proves difficult or inappropriate, a question can be prepared directly from the summary statement. First, determine if the summary statement can be restated in if-then form. If so, consider one of the following types of questions.

1. Given the antecedent, the student selects or supplies the consequent.

2. Given the consequent, the student selects or supplies the antecedent.

If all else fails, delete one or more elements from the summary statement. Turn the remainder into a question. Remember to paraphrase if possible. The question should contain no substantive words in common with the summary statement.

The correct answer to each question must be indicated. The answers to short answer items must be complete and, where applicable, alternative acceptable answers indicated. The wrong answers (distractors) should be provided for multiple choice items.

Selecting Item Format

Multiple choice and short answer items (items requiring a response of no more than a sentence or two) are desirable. No item format is definitely ruled out but true-false items and multiple response items containing only two response alternatives have serious limitations when used to measure comprehension. With either of these two item formats, the student has a fifty percent chance of guessing the correct answer even if he does not comprehend the material. Thus it takes a longer test to get a reliable indication of a student's comprehension of material. The true-false items have particularly severe limitations in that they call for an unqualified judgement of true or false. Since most important concepts in a discipline are neither unequivocally true nor false, a test item writer typically resorts to testing a student's knowledge of discrete factual information (Gronlund, 1968; Thorndike & Hagen, 1969).

Require discriminations as fine--but no finer than--the student could reasonably be expected to learn from the text. Don't split hairs. Whereever possible and appropriate, prepare items that involve pictures, graphs and schematic representations instead of just words and numbers.

Editing Questions

After the test items have been constructed, both the initial summary proposition and test items must be reviewed to help insure validity. The following questions may be used as a checklist in the review process.

1. Does the summary statement accurately reflect the main idea in the subsection?

2. Does the question faithfully reflect the summary statement?

3. If the question does not involve application to a new example, would it be possible to create a new question that does?

4. Does the question repeat substantive words from the summary statement? If so, could these reasonably be replaced with synonymous words or phrases?

5. Is the question simply and clearly written?

6. Does the question have any features that might cause a student to fail even though he has comprehended the material?

7. Does the item contain clues which might allow a student to answer correctly even though he has not comprehended or perhaps not even have read the material?

In summary, if text materials contain pertinent information and if students can read and understand that material, test items can be constructed which will facilitate the student's processing of information. Summary statements are constructed from the text material as an initial step. Then, items assessing the student's comprehension of the text are constructed from the summary statement. Finally, the items and the summary statement are reviewed and edited by both subject matter experts and item writers to improve validity and reliability.

Initial Development Effort

This section of the report discusses the application of the procedures outlined in the General Discussion section to a large enrollment economics

course. The item writing effort included both economists and educational psychologists in all the aspects of course development and, more specifically, question construction. Both an initial procedure and a modified procedure are discussed.

Unit of interest. Initial examination of the basic text used in the course, Economics (Ninth Edition) by Samuelson, indicated that constructing questions for each paragraph would be both unfeasible for the item writing team and counterproductive for students using the material. Subsequently, the decision was made to focus upon the chapter subsection as the unit of analysis. At least one summary statement was to be written for each of the subsections of the text, excluding introductory and summary subsections. A content specialist from the Economics Department at the University of Illinois analyzed each of the subsections and constructed the summary statements using the verbatim language of the text.

Included below is a summary statement constructed for material found on page 379 of the text. The statement was constructed on May 13 and then given to test item writers for their use in constructing the item.

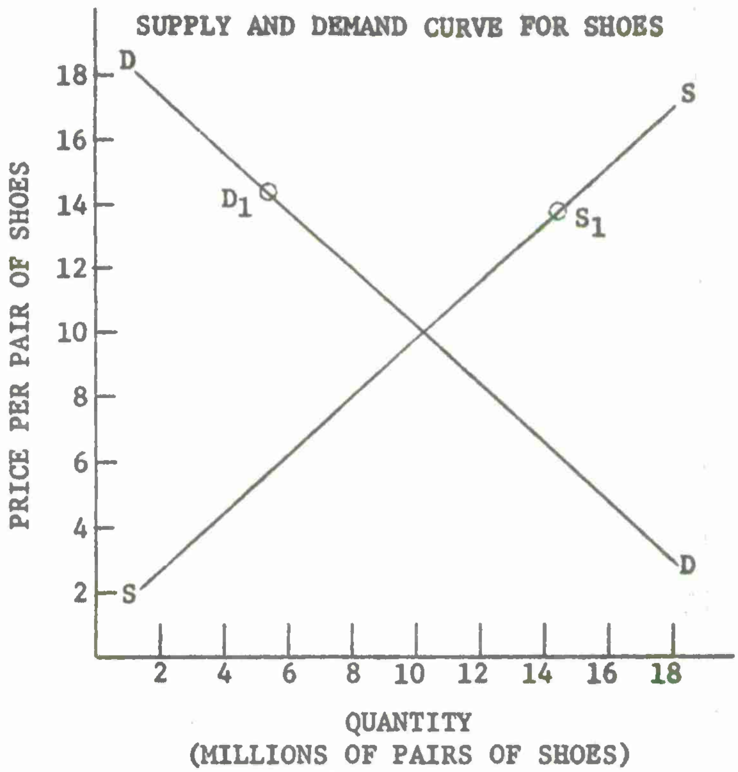
> Summary Statement: The downward sloping demand cruve and upward sloping supply curve intersect and determine an equilibrium price. Any deviation of price from this equilibrium level will create a situation (oversupply or excess demand) which will tend to force price back to the equilibrium level.

Item construction. As each chapter was analyzed and a summary proposition for each of the subsections written, graduate students in educational psychology began the task of constructing questions. Two item writers were initially employed in the construction of the test items, writing items for subsections in alternative chapters of the text. As one

item writer finished constructing the item for the subsections of a chapter, the other item writer would review and edit the items. After the items were reviewed and edited by the item writers they were reviewed and edited by other members of the item development staff.

For example, the item constructed initially to test the students' understanding of the above summary statement is shown below.

Test Item:



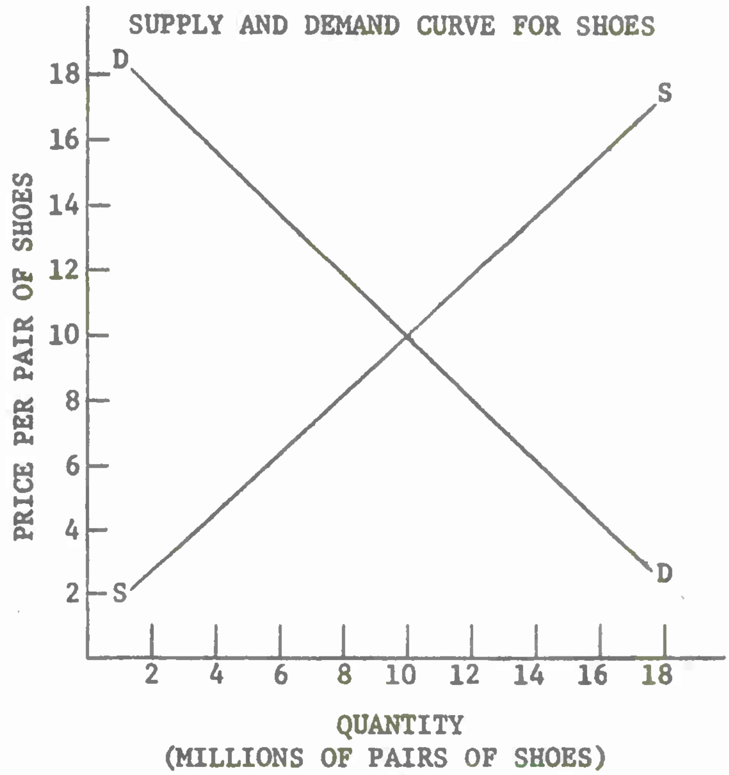
Depicted above is a demand and supply schedule for children's shoes. If the price for children's shoes were $14, one would say there is ...

1. an excess demand for shoes at a price of approximately $9 per pair.
2. an insufficient supply of shoes at a price of approximately $9 per pair.
3. an oversupply of shoes at a price of approximately $9 per pair.

Figure II.1 diagrams the entire process. (1) A subject matter expert summarized the main ideas in each chapter and wrote the summary statements. (2) Educational psychologists construct and (3) then edit the items. (4) The summary propositions and corresponding items were typed and (5) subsequently reviewed and edited by a curriculum developer. Items were then edited (6) by another subject matter expert and (7) retyped as needed. Finally, the items were reviewed (8) by two undergraduate students who read the text material and then answered the questions. When all was in order the items (9) were put onto PLATO. At each step in the procedure, the individual completing work on an item would date and initial the item to indicate when the work was completed and who did the work. Also, each item was edited by working directly on the original version of the item. If extensive revisions of an item were necessary, revised items were written on separate sheets of paper that were, in turn, attached to the initial work. At each stage of the process the development team had a complete copy of all revisions of an item as well as data indicating when work was being completed.

Question No. 20.2 was reviewed and edited throughout the process such that the final form of the item was as follows:

20.2



Depicted above is a demand and supply schedule for children's shoes. If the price of shoes were $14 per pair, one could justifiably infer that ...

1. inventories of shoes would increase.
2. inventories of shoes would decrease.
3. $14 per pair is the equilibrium price of children's shoes.

This is an exact duplication of what the student actually viewed when he sat down at the PLATO terminal.

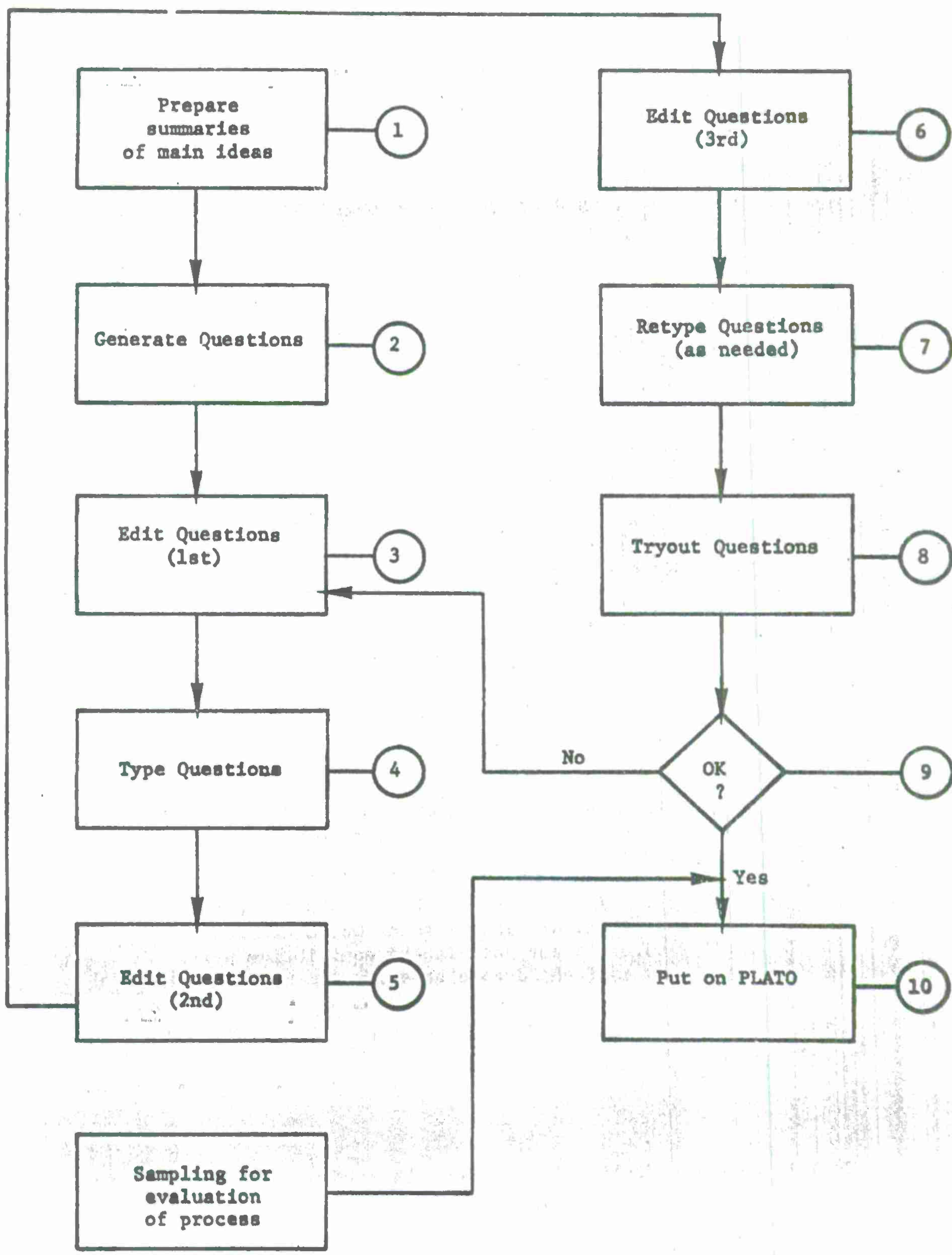


Figure II.1. The process for constructing questions.

Revision of the procedure. The data gathered early in the development effort was used to streamline the procedure as certain steps were discovered to be dysfunctional or redundant. Steps (4) and (5), typing and review by a curriculum developer, were eliminated. Typing was found to be unnecessary at the early stage of an item's development and involved large amounts of typing for little purpose. The editing of the item by the curriculum developer proved unnecessary since the initial editing by the item writers and the detailed scrutiny by the content specialists and students proved to be sufficient. The method of collecting data concerning the process proved to be cumbersome and time consuming but was used throughout the process. The procedure was modified for the second development effort and will be discussed later in this report.

Data section. Personnel requirements for the item construction process using the modified procedure discussed in the preceding paragraph included the part time efforts of two content specialists, two educational psychologists and one typist-stenographer for six months and the services of an additional item writer during the last month of development. There were two fairly serious problems with the process. The first stemmed from item writers' unfamiliarity with the subject matter. The second was the difficulty in arriving at a concensus between the subject matter experts and the item writers as to what constituted a "good" item.

Table II.1 lists the number of subsections of text, as well as the number and type of items constructed and the item format. From the data listed, it can be seen that of the 291 items constructed, approximately 55% were application items while 45% were paraphrase items. The predominant

Table II.1

Number of Subsections of Text, Item Format, and the Number and Type of Items Constructed During First Development Effort

Number of subsections in text	282
Number of items constructed	291
Type of item	
Application	162
Paraphrase	129
Format of item	
Multiple choice	259
Two choice response	7
Fill in or short answer	24
Matching	1

type of item was multiple choice with nearly 89 percent of the items using this format.

Evaluation of the Items

Face validity refers to whether or not a test item is perceived to measure what it is supposed to measure. For example, a test item designed to measure a student's comprehension of economics would have little face validity if it did not contain terminology appropriate for economics or if it did not test economic principles or concepts. This issue of face validity is important because the perceptions of the applicability/appropriateness of an item will influence the decision to either use the item (or remain involved in the development project) as well as influencing student affect toward the subject matter. If students perceive test items to be inappropriate or trivial they are less likely to have a positive affect toward either content in general or the specific course.

To measure face validity of test items is primarily a subjective process. The informed opinion of various experts and of the students actually using the material is the only evidence relevant to this point. Other sections of this report speak to this point.

In general, content validity is perhaps easier to insure than is face validity. The procedure used in constructing the items helped insure that the test items systematically and consistently paralleled the content of the text. Test items were constructed for all of the required subsections of the text. Every item was constructed so that students had to have comprehended the text material if they were to answer it correctly.

The construct validity of the items is more difficult to insure. Do these items actually measure comprehension of the content? The article by Anderson, "How to construct achievement tests to assess comprehension," appearing in Review of Educational Research, 1972, 42, 145-170, contains a justification for an emphasis on paraphrase and application questions.

Evaluation of the item development procedure. Using the procedure outlined earlier in this section, a total of 291 test items were constructed. Although this was a formidable achievement certain dysfunctional aspects of the procedure were encountered.

The coding of questions and the keying of the items to specific portions of text were awkward and not specific enough in some instances. If a specific subsection of the text was "meaty" and required a number of summary propositions, it proved difficult for the item writers to identify the pertinent portions of the subsections. During review and final editing of items, it was often necessary to go back to the text and examine the congruity between the content, the summary proposition and the question. Since two or sometimes three summary propositions were constructed for some subsections of the chapter it was difficult to quickly locate the appropriate content.

The record keeping system in which each item was initialled and dated at each stage of the process also proved to be awkward. At any given time data was available to allow us to monitor the development process for items or groups of items. But, to monitor the development process as a whole required transcribing data from the individual test items into log books and master flow charts. The time expended in paper work alone put

an additional strain on already limited resources. A simplified record keeping procedure was developed and successfully used in the second development effort described later in this section.

A serious oversight of the developers was the failure to provide both the item writers and instructor for the course with their own list of test items. As items were written, rewritten, edited and re-edited it was difficult to keep a close monitor on the items. The problem was compounded in that the course instructor needed access to the items so that he could monitor his students' progress and be prepared for any questions they might have concerning the items. This minor problem was quickly solved by making multiple copies at each stage.

The most serious problem encountered in the question writing phase of the project stemmed from the lack of a common orientation to the project on the part of both educational psychologist item writers and the subject matter experts. Due to a lack of understanding of each others' respective fields, communication between the subject matter experts and the item writers was sometimes difficult. In a number of instances items were edited and re-edited as many as six times because of what appeared to be "trivial" points. For example, an item would be constructed and edited by the item writers. The item would then be sent to the subject matter expert for review. The subject matter expert would rewrite the item and offer what he felt was a more acceptable item and then send it back to the item writers for review. The item writers would review the item and make what they felt were essential corrections or modifications and would then route the item back to the subject matter expert.

He reviewed the item and would feel that meaning of the item was changed and would correct the correction. This correction would be reviewed by the item writers who, in turn, would make corrections. The cycle would be repeated and repeated until concensus would finally be reached (as well as ulcers activated by all concerned) or the item was deposited in a separate file to be reconsidered at a later time. Recommendations to preclude this type of situation will be presented later in this section.

Second Development Effort

This section discusses the procedures used in constructing questions to be used by CAISMS in conjunction with a basic American history course taught at a junior college. Many of the shortcomings and dysfunctional aspects of the procedure initially used were corrected in this second effort.

Roles and tasks assigned to members of development team. Construction of the questions to be used in the history course began in September 1973 with a greatly streamlined staff. The test item writer was a graduate student in educational psychology and the subject matter expert was the history instructor for the course. These two persons constructed all the summary statements and questions to be used in the course during the last five months of 1973. Clerical assistance was provided by a typist-stenographer from the Department of Educational Psychology. General supervision of the development effort also came from the Department of Educational Psychology.

The abbreviated team possessed several advantages over the team that constructed the questions for the economics course. The subject matter

expert who constructed the summary propositions for the history course was also the subject matter expert who reviewed and edited the items. All decisions concerning the content of the course were made by one person thereby alleviating any necessity for compromise, discussion or consultation with other historians who might be affected by CAISMS. The graduate student in educational psychology that constructed the test items had taken his undergraduate degree in history and already possessed a basic understanding of the subject matter. Therefore it was possible for him to write items that needed much less revision by the subject matter expert.

Development of summary statements and questions. The general procedure outlined in Section I was followed in the development of the summary propositions and questions. However, the process was structured so as to preclude some of the problems encountered in the initial development effort.

The units of analysis for the history course were the subsections of the individual chapters of the text, The American Nation (Second Edition) by Garraty. To insure standardization of numbering and codification, the 97 subsections of the text (only the first seven chapters of the text) were listed sequentially with an identifying code number. Page references for each subsection were also listed. The decision was made to construct only one summary statement for each chapter subsection. Thus, before summary statements or items were constructed, the development team knew exactly how many summary statements and items were to be written, the code numbers for each of the 97 items and the page references.

Figure II.2 depicts the subsection breakdown completed prior to actually constructing summary propositions and questions. As can be seen,

1. The Age of Discovery and Settlement

1.1	Columbus and the Discovery of America	15-21
1.2	The Indian and the European	21-23
1.3	The Spanish Decline	23-24
1.4	English Beginnings in America	24-27
1.5	The Settlement of Virginia	27-30
1.6	The Pilgrims	30-31
1.7	Massachusetts Bay Puritans	31-33
1.8	Other New England Colonies	33-33
1.9	French and Dutch Settlements	33-35
1.10	Maryland and the Carolinas	35-37
1.11	The Middle Colonies	37-39
1.12	The Settlement of Georgia	39-40

Figure II.2. Subsection breakdown for Chapter 1 of the text, The American Nation, indicating the subsection title, code number and page references.

chapter subsections have been identified by code number, subsection title and appropriate text pages.

A firm time schedule for the completion of tasks was decided upon after taking into account the nuances of the school calendars. Work on the construction of summary propositions began the week of September 14.

The record keeping procedure was also streamlined in the second development effort. During the initial project summary statements and questions were constructed chapter by chapter with data recorded directly onto the summary statement and question. During the second effort, manilla envelopes were prepared which listed each step in the procedure; space was provided for initialling and dating as each stage was completed (see Figure II.3). When the subject matter expert began constructing summary propositions she wrote:

1. the identifying code number,
2. subsection title, and
3. page reference numbers

on the top of a sheet of paper and wrote the summary statement directly below the identifying information. When summary statements were constructed for all the subsections of a chapter they were placed in the envelope and appropriately initialled, dated, and given to the item writer who then constructed the questions. When finished with this phase of the project, the item writer simply returned the summary statements and items to their envelope, initialled and dated it, and sent them to the subject matter expert for review. For example, during the week ending September 26, the subject matter specialist finished constructing the summary statements for

The American Revolution	Chapter 1	
	Date	**Initial**
Summary statements written	26 Sept	BF
Items prepared	5 Oct	EJW
Items edited by content specialist	8 Oct	BF
Items reviewed by test specialist	8 Oct	EJW
Differences resolved by C.S. and T.S.	8 Oct	EJW,BF
Items typed	10 Oct	BA
Items reviewed by student		
Items put in PLATO system		

Figure II.3. The record keeping procedure used for the development of test items for Chapter 1 of *The American Nation*.

material contained in Chapter 4 of the text. Figure II.4 depicts the identifying code number, subsection title, page reference numbers and actual summary statement for this section of text.

The item writer wrote the following question to assess students' comprehension of the material.

> If American forces were to defeat Japanese forces at Tarawa in 1943, <u>in the same manner</u> that colonial forces defeated British forces at Yorktown in 1781, what would be essential for an American victory?
>
> 1. Control of the sea around Tarawa
> 2. Larger numbers of trained troops
> 3. A more efficient supply system
> 4. A more powerful navy
> 5. Superior artillery

After reviewing the item, the subject matter specialist felt that the item would be extremely difficult for her students and suggested that an entirely new item be constructed. After discussion with the subject matter expert, the item writer concurred with her opinion and constructed the new item shown below.

> Assume that you are a British tactician advising General Cornwallis during the Battle for Yorktown. You are very worried about the possibility of a British defeat because of certain information you have just received and want to warn Cornwallis. What would you be warning Cornwallis about?
>
> 1. The superior logistical system of the Colonial and French forces
> 2. The numerically superior Colonial and French forces
> 3. The French fleet's control of Chesapeake Bay
> 4. The Colonial and French possession of superior artillery

The subject matter expert, in turn, reviewed the item and considered it fully acceptable.

4.9 Victory at Yorktown 172-173

Everyone realized that the rebellion could not be suppressed unless the British could maintain control of the Atlantic; [and when Cornwallis surrendered at Yorktown, the war was over for all practical purposes.]

Figure II.4. Identifying code number, subsection title, page reference numbers and summary statement.

This procedure allowed the review, editing, summary statement construction and question writing for a number of chapters concurrently with a minimum of confusion. For example, during the week of October 19, 1973, the following tasks were completed by the team:

Subject Matter Expert	Item Writer
Preparing summary propositions for Chapter 6	Preparing items for Chapter 5
Editing items prepared for Chapter 4	Editing items for Chapter 3
Re-editing items prepared for Chapter 3	Re-editing items for Chapter 2

Meanwhile, the typist-stenographer was preparing the final typewritten copies of materials for Chapter 1.

Data section. Personnel and time requirements have been discussed earlier in this section. Table II.2 summarizes the work that was accomplished.

Evaluation of the items and the item development procedure. The issues of face validity and construct validity have been discussed earlier in this report. Content validity stems from the fact that 95 summary propositions and questions were developed for the 97 subsections of text. Propositions and questions were not developed for two of the subsections of text because, in the opinion of the subject matter expert, the issues discussed were trivial.

The procedure used to develop the items proved to be efficient and workable. Minimal expenditures of resources were required for the development of apparently satisfactory items. A detailed evaluation of the junior college project will begin later this year which should provide additional data about the suitability of the items.

Table II.2

Number of Subsections of Text, Item Format, and the Number and Type of Items Constructed During Second Development Effort

Number of subsections in text	97
Number of items prepared	95
Format of item	
Multiple choice	85
Two choice response	4
Fill in or short answer	6
Type of item	
Application	26
Paraphrase	69

Recommendations

Based upon the experiences gained in the two development efforts described thus far, the procedure to develop questions for maintaining the attentive processing of text material outlined below is recommended.

1. The project director, in consultation with the item writers and the subject matter experts, must decide upon an appropriate and feasible unit of analysis.
2. Once the unit of analysis has been decided upon, assign each unit a:
    a. descriptive title
    b. identifying code number
    c. page reference number
3. List the units and identifying titles/numbers and decide upon intermediate target dates for the various tasks to be performed. If circumstances dictate, these target dates may be changed but they initially provide clear objectives for members of the development team.
4. A minimum of three copies of the target dates and list of the units of analysis should be prepared. The director of the project receives one copy as does the item writer and subject-matter expert. Note, if more than one item writer or subject matter expert is used <u>all</u> members should receive a copy of this material.
5. After this preliminary work has been done and when both subject matter experts and item writers are familiar with the general procedure outlined in the General Discussion section of this report, actually begin work.

6. Record keeping is essential for formative and summative evaluation of the item construction phase of the entire project. The procedures discussed in the Second Development Effort section and the general procedure described in the Recommendations section have been proven to be both effective and useful in the construction of test items.

References

Anderson, R. C. How to construct achievement tests to assess comprehension. *Review of Educational Research*, 1972, *42*, 145-170.

Bormuth, J. R. *On the theory of achievement test items*. Chicago: University of Chicago Press, 1970.

Gronlund, N. E. *Constructing achievement tests*. Princeton, N. J.: Prentice-Hall, 1968. Pp. 44-45.

Thorndike, R. L. & Hagen, E. *Measurement and evaluation in psychology and education* (3rd Edition). New York: John Wiley & Sons, 1969. Pp. 96-97.

III. THE DEVELOPMENT OF A COMPUTER PROGRAM TO DELIVER CAISMS

W. Barry Biddle and Thomas H. Anderson

The initial step in the development of the delivery system was an analysis of the basic requirements of a study management system. Five major areas were identified as necessary elements of such a delivery system.

Basic Requirements of the Study Management System

First, the system must be capable of presenting questions to the students. This entails storage of a large number of questions in some accessible form so that the appropriate question can be given to each student at the right time. Second, the delivery system must then be able to accept in some manner the student's response and judge it. Accepting and judging student responses requires the delivery system to be an interactive system. Judging the response also requires the system to store all of the correct answers in some form which allows them to be matched to the corresponding question and then compared to the student's answer.

A third major requirement placed on the delivery system is record keeping. The two basic record keeping needs are: 1) to record and keep track of a student's present position in the course, and 2) to store the student's responses and an index of whether they were judged correct or not. The fourth area identified as a necessary element of the delivery system was feedback. Some means of informing the student of his progress and performance would have to be included in the delivery system. Remediation was the fifth area of concern identified. It seemed that if a study management system were to function properly a system of dealing

with students not meeting the prescribed criterion would have to be developed. For most efficient performance of this task an automated procedure activated and controlled by the delivery system seemed desirable.

The PLATO IV hardware system at the University of Illinois was selected to implement the study management system. In addition to the fact that it is convenient to the researcher the system appeared to have all the necessary capabilities to handle the project.

The Tutor Program

Development of the TUTOR programs (PLATO IV's language) to implement the study management system began with the design of a component to display questions and accept and judge answers. The design of PLATO and the TUTOR language seemed well suited for this aspect of the study management system. The language has many commands which simplify the creation of visual displays. TUTOR also performs many answer judging routines automatically or at the direction of a single command. At this point some additional characteristics which seemed desirable were added to the list of considerations for the displaying and judging component. In order to increase the flexibility of the study management system the delivery program should be capable of handling as many different forms of questions and answers as possible. This made it desirable to have a delivery program which would be capable of displaying and judging many different types of questions. Development of the capability for handling adjunct materials such as graphs, charts, slides and sound presentations would also create flexibility. For purposes of efficiency, consideration was given to constructing the program so that a non-programmer could enter,

delete or change questions. The final consideration was that the items be given in a different random order to each student.

The driver program. Given the requirements and considerations it was decided to adopt the concept of a "driver" program and various storage files. A driver program repeats a series of operations over and over using different values or materials each time. This seemed to be appropriate for the study management system, since the basic design is repeated presentations of questions. The process not only repeats itself for each quiz, but also for each item within a quiz. The files are simply addressed lists of information, or in other words lists of information which have unique identifiers attached to each individual piece of information. Using a file system of this type, questions can be entered and edited by a non-programmer since it requires nothing more than normal typing skills.

The system of files. Three basic sets of files were developed. One file simply contains the page numbers for each reading assignment. The other files are the question and answer files. The answer file contains the correct answer and any acceptable alternatives. The unique identifier for each answer is matched to the question identifier with a prefix to indicate that it is an answer (e.g., A120149). This enables easy editing of the files. The question files contain all of the information that is displayed for the student in connection with each question. The unique identifier for questions is composed of the prefix Q, indicating that it is a question followed by a 6-digit number. The first three digits represent the page in the text book where the information relating to the question can be found. The second three digits represent

a sequential numbering of the questions as they were written (Q120149). Adjunct materials such as graphs and charts are stored in a separate section of the question files. The identifiers for the adjunct material are also keyed to the question number, but preceded by a letter which indicates that they are adjunct material (e.g., G120149). The driver program accesses all of these files and also calculates and stores information in a record component.

Coordination between the driver and files. The driver program begins by accessing the assignment file and displaying on the student's screen the next assignment. A counter is used to keep track of how many assignments the student has completed. The value of this counter determines which assignment the student will receive and which set of questions the student will be given. After the student has been given an assignment a flag is set in the driver program which indicates that this student has received his assignment. The driver program then automatically signs that student off the system. The student will be kept off of the system for seven minutes. This time is designated as reading time for the assignment (see Figure III.1, parts 2, 2a, and 2b).

When the student next signs on to the system the driver program will check to insure that the student has in fact been off for seven minutes or more. If the student has waited the proper time he will be allowed to take the quiz for that assignment. The driver program will then access the appropriate question file and display the material for the first item. If a marker in the question file indicates that there is adjunct material for this item the driver program will also access the file containing the

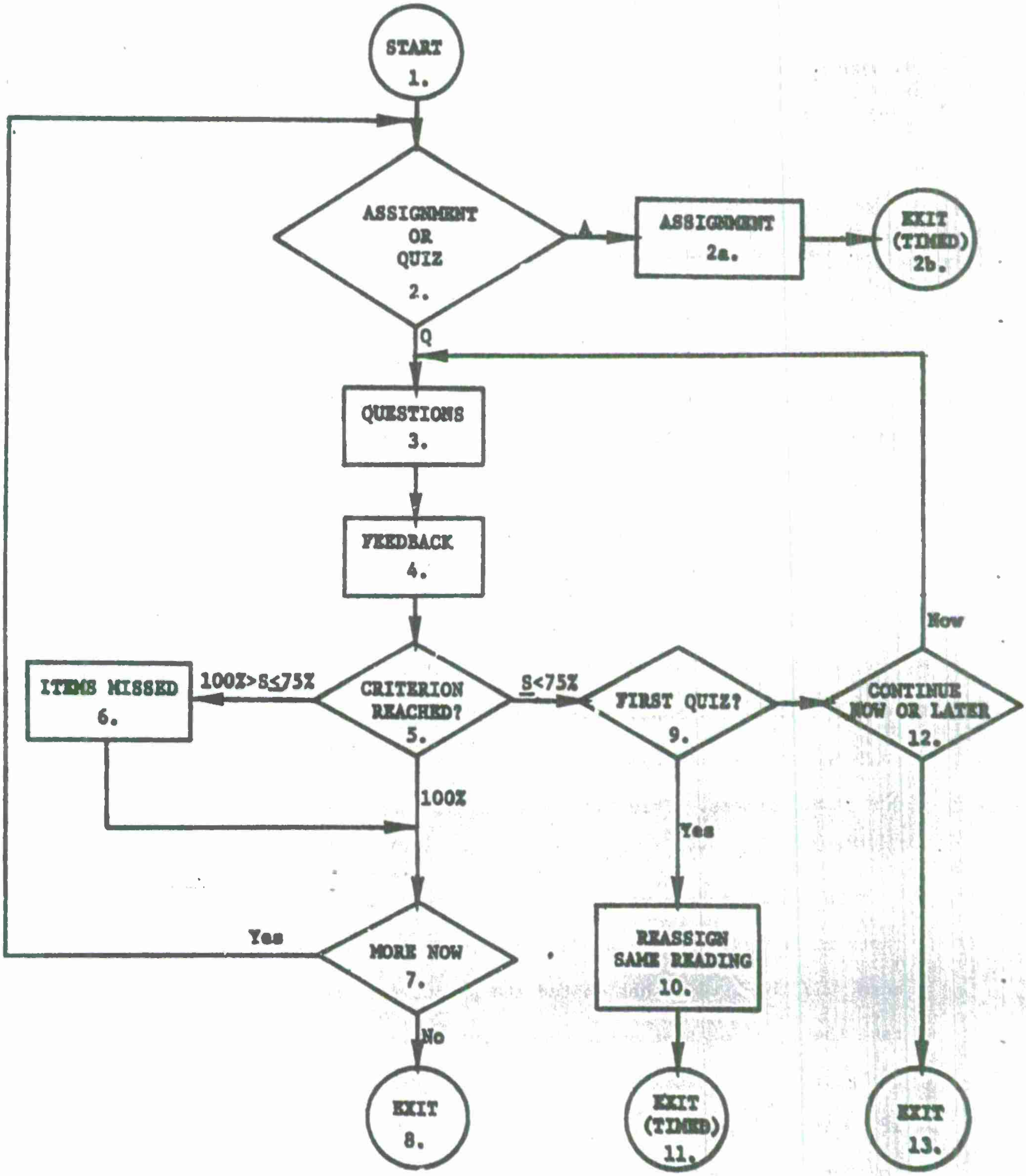


Figure III.1. A flow diagram and description of the computer based instructional process used in CAISMS.

1. <u>Start</u> - When students sign-on to the CAISMS system the time since they last received an assignment [Exit (timed) 26,11] is checked to make sure they have waited the suggested seven minutes for reading the assignment.

2. <u>Assignment or Quiz</u> - This branch determines whether the student should be given a new assignment or a quiz.

2a. <u>Assignment</u> - If the student is to receive an assignment a message is displayed giving the appropriate page numbers.

2b. <u>Exit (Timed)</u> - After receiving an assignment the student is automatically signed-off the system and a timer is set to be used for signing-on in "Start".

3. <u>Questions</u> - When the student is to receive a quiz, the first thing that is determined is which quiz--original or retest. Then the items for that quiz are displayed in random order. The students' responses are judged and the results are stored.

4. <u>Feedback</u> - When a student has completed a quiz a message is displayed showing the percentage of items he answered correctly.

5. <u>Criterion</u> - Criterion for moving on to the next assignment is 75%. After the student has been informed of his performance, one of three subroutines is begun based on whether or not he reached criterion.

6. <u>Items Missed</u> - If the student reaches criterion, but does not answer all of the questions correctly he is given the items he missed over again. Each item remains on the screen until he is able to answer it correctly.

7. <u>More Now</u> - Either after the student has retaken the items he missed or if he answered all questions correctly on the original quiz, he is given the opportunity to get a new assignment or to sign off.

8. <u>Exit</u> - If the student chooses to sign-off at this point all of his records are updated and he leaves the system.

9. <u>First Quiz</u> - If the student fails to reach criterion the decision as to what happens next depends on whether or not it is his original quiz or a retest.

10. <u>Reassign Same Reading</u> - If it is the first quiz, a message is displayed saying that the student should reread the assignment. The page numbers are repeated in this display.

11. <u>Exit (Timed)</u> - After being told to reread the student is automatically signed off the system and the timer set to prevent the student from signing-on for seven minutes.

Figure III.1 (continued)

12. <u>Continue Now or Later</u> - If the student fails to reach criterion and it is a retest he is given the choice between retaking the test immediately or at a later time. If he chooses immediately he is given the quiz again.

13. <u>Exit</u> - If the student chooses not to take the retest immediately he simply signs-off the system.

Figure III.1 (continued)

appropriate information and present it. Next the driver program indicates to the student that he is to record his answer. When the student has entered an answer and indicates he is ready to have it judged the driver program accesses the appropriate answer file and obtains the stored correct answer and its acceptable alternatives. If the student's answer matches one of the stored answers the driver program puts a correct marker in the student's records. If he is wrong an incorrect marker is entered into the records. (Records will be discussed in more detail in a later section.) This process is repeated for each item of the quiz (see Figure III.1, part 3).

When the quiz is completed the driver program displays a feedback message and provides remedial work if necessary (feedback and remediation will be discussed in more detail later). When the student is ready for the next assignment the entire process begins again (Figure III.1, parts 5-13).

Problem with TUTOR program. The present conception and design of the driver program and files has led to some problems. It has not provided as much flexibility in presenting various question forms as would ultimately be desirable. The present design allows one response to be entered, judged and scored for each display. When judging is complete the display is erased. This prevents the use of questions requiring several different responses such as multiple-blank fill-ins or matching items.

Another problem which has arisen is related to the size of the files. The large files that have been created prevent the simultaneous access of

all of the assignments at any one time. This problem is an interaction of hardware limitation and software design. The computer memory allocation for the CAISMS system is smaller than the total space needed to store all of the files. This has not been a prohibitive problem since generally students can access most portions of the material that they are working on when they want to. Additional consideration, however, needs to be given to this problem.

Feedback and Remediation

The design of the computer assisted study management system requires that some form of feedback and remediation be given to the students in order to adequately manage student processing. In keeping with the simplicity of the concept of the study management system, a simple feedback message was designed. After a student has finished all of the questions in a quiz he receives the feedback message which indicates what percentage of the questions he answered correctly. The driver program computes this percentage and controls presentation of the message.

Rereading the text book and retaking quizzes were chosen as the basic means of remediation for CAISMS. In terms of the development of the delivery system this meant that questions had to be re-shown and new student answers judged, all without destroying or altering data collected on the original quiz. To handle this problem the judging and storing routines of the driver program were made conditional so that for the original quiz one method of judging and storage would be used while for retakes other methods and storage locations would be used.

Three basic remedial routines were developed. The decision as to which one a student will receive is determined by his performance on the preceding quiz. Seventy-five percent correct was established as the criterion for moving on to the next assignment. One of the feedback-remediation routines is administered to students who reach this criterion, but who fail to answer all questions correctly. This routine displays, one at a time, those questions that the student answered incorrectly. Each question remains on the screen until it is answered correctly. Again the driver program is responsible for displaying these items. It determines which items were answered incorrectly by searching its data store for that individual looking for incorrect flags under the question numbers of the present quiz.

If the student fails to reach criterion on the first quiz he is asked to reread the same section of the textbook and then he is automatically signed off the system. He is not allowed to sign back on until seven minutes have elapsed. At this time he is given the same quiz over again. If the student fails to reach criterion on the second retest he is given the quiz again. This time no delay is enforced, however this cycle will continue until the student reaches criterion. It is hoped that this mechanism will encourage the student to process the text material.

All quiz questions are displayed by the same routine in the driver program with only the judging and storage routines changed for different quizzes. This method was found to be more efficient particularly in terms of computer memory space than an earlier alternative program which used separate displaying and judging routines for different quizzes. The earlier program was also limited to one retest.

Student Records and Data Storage

Another requirement of CAISMS is that several different types of data be kept in order to monitor student progress and achievement. Some of the data that it was thought would be desirable to record was student progress (which quiz or assignment he is on), how well he is doing, how he answers the questions, and when he last used the CAISMS system for study.

Much of the desired record keeping and data storage could be handled by existing PLATO routines and systems programs. For example, the "student record" system of PLATO was designed to handle record keeping for individual students. Student records provides 150 variables or storage locations for each student listed in the course. Various forms of data can be entered into these variables from TUTOR programs on PLATO. Student records is convenient for CAISMS since the variables in student records can be set, changed and used in the driver program. This allows the use of the same variables for storing data and controlling the driver program. In this manner one variable is used to record the student's present position in the course. This same variable controls, via the driver program, which questions the student will see. Also all of the conditional routines in the driver program, such as the judging and storage routines for different quizzes, are controlled by values which are stored in student records for individual students. This way when a student signs on his individual records are used to supply the values needed to operate the driver program.

"Student records" also automatically record the last time each student has used PLATO, therefore an updated record of the last time each student has used the study management system is automatically provided. "Student

records" also keeps updated records of the length of time the student has been on PLATO and the number of different days that he has been on in addition to the number of individual times the student has signed on to the system.

Another PLATO system subroutine provides a simplified means of storing data on student responses. The subroutine (Flags) used stores a 1 for questions answered correctly and a 0 for questions answered incorrectly; at present this is the only item information that is retained. Due to space limitations storage of actual student answers is not possible at this time. The TUTOR programming involved in this process is relatively straight forward and could be easily added when space constraints are relaxed.

The one problem that does exist with the present record keeping system is that the student record system of PLATO does not allow certain manipulations of the data which are desirable in a project such as CAISMS. At present the stored records for students are accessible only on a one student at a time basis. This prevents summing data across students or across quizzes on individual questions. Since student record data cannot be manipulated easily on PLATO, some information such as item analyses and average class performance is not readily available. Presently, this portion of record keeping and data manipulation is accomplished by having a clerk copy the relevant information for each student from the PLATO records system and entering this information on the IBM 360 computer system. This system is then utilized to produce summary data.

Summary

The necessary elements of a delivery system were determined to be a means of displaying items, judging student responses and storing the responses. A delivery system designed to handle these needs was developed. The delivery system is based on a driver program which accesses a series of files. The files consist of assignments, questions, answers and other adjunct materials. With this delivery system a user can implement a study management system in almost any discipline that he desires. To implement a new course using this system, the user must have the items he wants to use and the correct answers. The user must also group the questions into clusters which correspond to the study assignments he wishes to make. This information must then be loaded into the delivery system files. A few changes in the driver program are also necessary to match up the labels of the new information with the labels used by the program to access the files.

IV. EVALUATION OF CAISMS

John R. Surber and Thomas H. Anderson

The evaluation of the first semester of the study management system proceeded along several lines. Data was collected that describe the student population enrolled in the course. Daily records were kept on student interaction with the terminals as a means of monitoring student study behavior. Student performance was measured by their scores on the PLATO quizzes as well as three hourly in-class exams. Finally student attitudes towards the course were obtained by means of an attitude questionnaire. These different types of data are summarized in Table IV.1. In addition a canonical correlation analysis was performed on all of the data in order to determine what variables best accounted for or were the best predictors of student performance.

Student Descriptive Data

The intent as stated in the proposal for this project was to test the study management system in a course which attracted students with a wide range of individual differences. Table IV.1 presents data describing the level of academic achievement, the year in school and area of curriculum of the students who participated in the first semester of the study management system. From these data it appears that the students did indeed have fairly wide ranges of ability and a broad interest pattern.

Student Study Behavior

The record keeping capabilities of the system allowed collection of data on student study habits such as the total number of days a student

Table IV.1

Some Descriptive Statistics of the Sample of Economics Students

	Percent	Mean	Standard Deviation
Class distribution			
Freshman	8		
Sophomore	39		
Junior	35		
Senior	18		
College curriculum distribution			
Agriculture	24		
Engineering	29		
Journalism	4		
Liberal Arts	41		
Physical Education	2		
Achievement			
High school percentile rank		81.71	18.10
College grade point average		3.94	.60
Study behaviors			
Number of days on system		24.75	9.88
Duration of study time (minutes per day on system)		24.0	8.8
Total number of hours on system		9.21	3.50

used the system during the semester, the amount of time the student was actually working on PLATO for each day that he used PLATO and the total number of hours the student used PLATO for the semester. Averages of these figures for all students are given in Table IV.1.

It was hoped that one of the effects of CAISMS would be to encourage students to study at a more even pace throughout the semester rather than "cramming" before exams. Daily records were kept on the number of item clusters completed by each student. Fifteen students were randomly selected and the mean number of item clusters completed each week was computed. These data are plotted in Figure IV.1. Class examinations were given in the sixth, twelfth and sixteenth week of the semester. It is evident from the graph that though progress was steady up to the first exam, there is a plateau after both the first and second exams. This scalloping effect indicates that CAISMS was not as effective in maintaining a steady rate of study behavior as we had hoped.

Student Performance on System

The criterion for passing a quiz on PLATO was set at 75%. The figures for percent correct on the first quiz show that on the average the majority of the students did not reach criterion on the first quiz. In fact only 18% of the students averaged criterion or higher on their first attempt at each item cluster.

Student Attitudes Toward the Course

In addition to data describing the student population and student behaviors a questionnaire was administered in the last week of the course

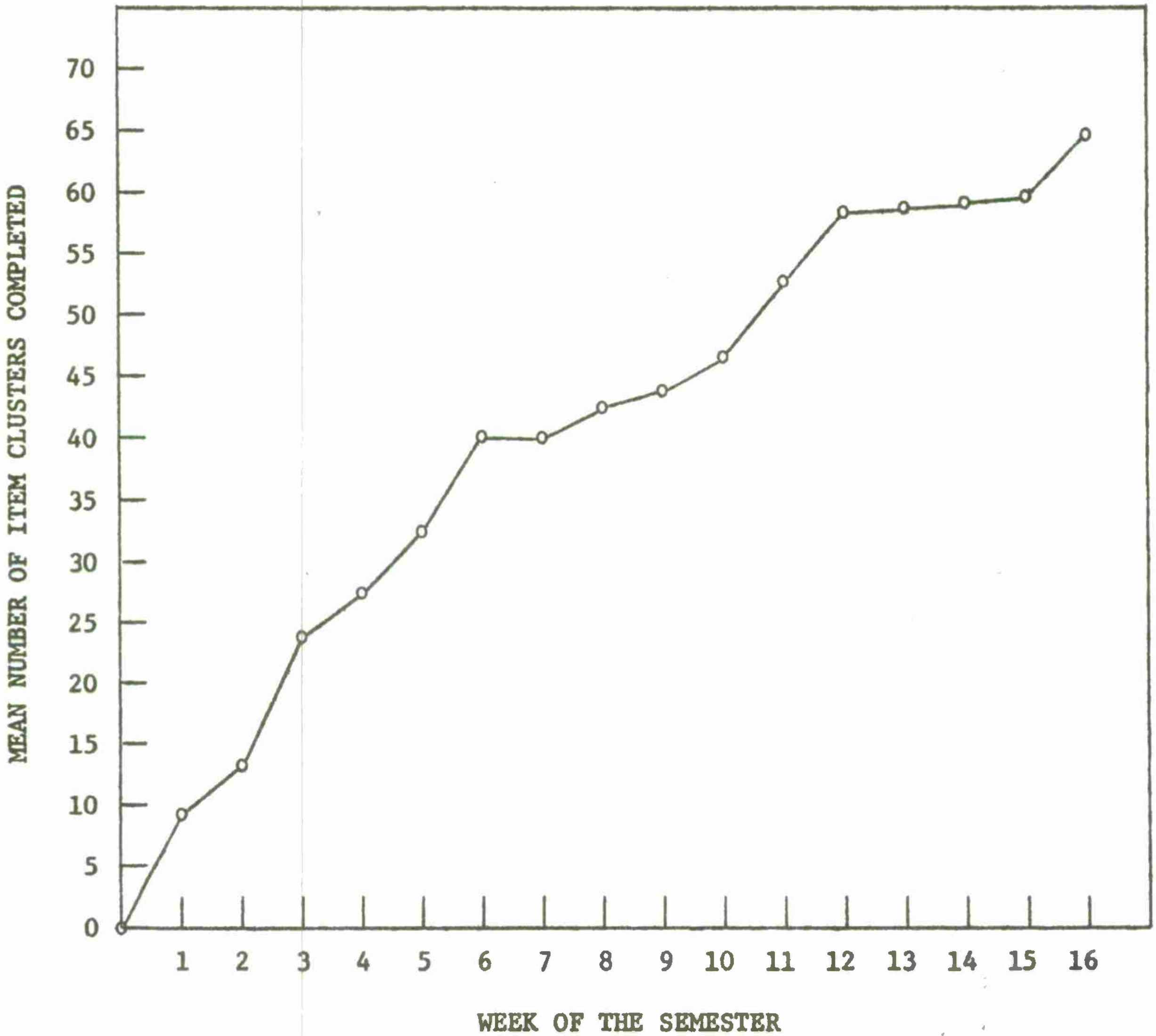


Figure IV.1. A cumulative study progress record of fifteen economics students.

to determine student attitudes towards various aspects of the course. These data are summarized below in a question and answer format.

<u>How did the study management system affect the study habits of the students?</u>

About 65% of the students felt that they spent more time studying for this course and about 50% felt that the system caused them to study more regularly. While only half of the students said that they studied more frequently in the library 65% said that the library was not a convenient place to study. Some students expressed very strong negative feelings about the library as exemplified by the following statement from a student:

> "It was a headache to go to undergrad library all the time when I needed to study ... also very crowded & disturbing in the undergrad library - hard to think."

<u>What effect did the students feel that CAISMS had on their course performance?</u>

About 80% felt that CAISMS improved their performance on exams. However, when asked to evaluate the course as a whole only 37% felt they learned more than in other courses.

<u>Were there aspects of CAISMS that the students found enjoyable?</u>

About 73% of the students said that interacting with other students around the terminals was either helpful or enjoyable.

<u>What other aspects of the system did the students like?</u>

Sixty-three percent of the students felt that the questions on PLATO helped them to know how well they were doing in the course. Students also felt that the questions were effective in helping them to review for tests. One of the more positive students summarized the benefits of the system in saying:

"... I felt that with PLATO I received more out of studying than I would have had I been left to do it alone."

<u>What did the students dislike about the system?</u>

Twenty-eight percent reported that one or more times per week terminals were unavailable and 66% reported that they encountered mechanical difficulties with the terminals two or more times a week. As one student expressed it:

"There were far too many mechanical difficulties with the system (i.e. 'memory allocation exceeded') that caused me to waste a large amount of time that I would have used to study for another course."

Although extremely negative attitudes were rare, one student summarized his feelings about the PLATO system with:

"... PLATO is unnecessary, expensive, useless, inconvenient."

<u>How would you characterize the overall attitude of the students toward the course?</u>

When asked if they would take another course taught the same way 61% of the students said that they would. Since the bulk of student complaints concerned mechanical difficulties of the system it is hoped that as these difficulties are ironed out that general student attitude will be even more positive.

<u>Regression Analysis</u>

Using the three class exams as three criterion variables and the eleven variables listed in Table IV.2 as predictor variables a multiple regression analysis was performed. It was encouraging to see that performance on the PLATO quizzes was one of the best predictors of performance on the exams. However, variables 10 and 11 were somewhat disappointing. If a student did

Table IV.2

A Listing of Eleven Predictor Variables and Regression Coefficients Used in a Significant Canonical Regression Equation with Three Exam Scores as Criteria

	Predictor Variables	Standardized Regression Coefficients
1.	Percent correct on PLATO quizzes	.64
2.	High school percentile rank	.34
3.	Sex of student	-.32
4.	Number of days on system per semester	.22
5.	Total hours on PLATO	-.22
6.	Cumulative hours of college credit	-.21
7.	Number hours per day on PLATO	.20
8.	Total hours of study for course per week	-.10
9.	College grade point average	-.09
10.	Reread assignment if fail first quiz	-.06
11.	Answered PLATO quizzes from memory	-.07

not reach criterion on his first attempt on a PLATO quiz he was instructed to reread the assignment and was prevented from getting back on the system to retake the quiz for seven minutes. Data from the questionnaire indicates that many students did not reread the assignment during the period they were locked off the system. A number of students also indicated that they answered questions with the aid of books and notes despite the fact that they were instructed not to. However, judging from the regression coefficients for variables 10 and 11 it seemed to make little difference in exam performance whether students followed or ignored these two instructions.

The fact that sex is a relatively good predictor of performance can probably be accounted for by the fact that many of the male students were in the engineering curriculum while many of the females are in home economics. The lower than expected relationship between GPA and performance may also be confounded with curriculum. Because the engineering curriculum tends to be more difficult than other curricula an engineering student with a relatively low GPA might perform quite satisfactorily in economics whereas a student in physical education might find economics very difficult. This seems to be born out by the fact that rank in high school was a relatively good predictor of performance.

Comparative Achievement

Although a comparative study of the effect of CAISMS was not planned for the first semester an informal comparison of final exam scores among the CAISMS section and five other sections of economics was possible since all six sections used the same text and took an identical final exam.

Although students in the CAISMS section scored higher than the other five sections the differences in mean exam scores were quantitatively small with 7% being the largest difference.

This comparison does not meet the requirements of experimental evidence since no effort was made to randomly assign students or instructors to course sections. It is interesting to note, however, that CAISMS appears to support the instructional program of the economics course and, perhaps, under controlled conditions will prove to effect a statistically significant difference in student performance.

V. TOWARDS A STUDENT RECORD KEEPING SYSTEM

Stephen M. Alessi

First Semester Procedures

The primary function of the CAISMS record keeping and data analysis system are to provide instructors and evaluators with summary data related to student progress and course material adequacy. The data were stored on the system basically as originally planned. In each student's personal file we kept track of a variety of variables including the last time he/she was on the system, total time on system, number of times on system, number of assignments completed, attitude information, and for each question encountered whether the student's response was judged to be correct or incorrect.

The above data were collected from the PLATO system sites and organized in two ways. First, a log book was used to record daily the first four of the above variables for each student. Second, a clerk recorded dichotomous item data for keypunching. Data in the log book were not subject to automatic computation. Dichotomous item data were analyzed via Fortran IV programs on an IBM 360/75J.

The analysis plans had two foci: 1) concerning individual students (Figure V.1) and 2) concerning assignments (Figure V.2). The student analysis initially included the student's ID, number of assignments completed, and for each assignment (called a cluster) completed, the number of items and what percentage of items was answered correctly.

```
                    COMPUTER AIDED INSTRUCTION STUDY MANAGEMENT SYSTEM                  ITEM
                    DEPARTMENTS OF EDUCATIONAL PSYCHOLOGY AND ECONOMICS                 ANALYSIS
                                  UNIVERSITY OF ILLINOIS                                PAGE 2
                                       19 FEB 74

ITEM CLUSTER 2

   83 STUDENTS HAVE COMPLETED THE ITEMS IN THIS CLUSTER

   INDEX OF DIFFICULTY
      MEAN= 2.9 QUESTIONS CORRECT
      SD= 0.84

   PROPORTIONAL INDEX OF DIFFICULTY (NUMBER CORRECT/TOTAL NUMBER ATTEMPTED)
      MEAN= 0.73
      SD= 0.21

DISTRIBUTION OF STUDENT PERFORMANCE

   NUMBER QUESTIONS       NUMBER            PERCENT OF STUDENTS
   ANSWERED CORRECTLY     STUDENTS
                                              0%,,,,,,,,,,,25%,,,,,,,,,,50%,,,,,,,,,,,75%,,,,,,,,,100%
           0                 0             0   *
           1                 5             6   ***
           2                17            20   **********
           3                40            48   ************************
           4                21            25   *************

INDIVIDUAL QUESTIONS DATA

                       NUMBER STUDENTS      PERCENT STUDENTS ANSWERED CORRECTLY
                       ANSWERED CORRECTLY
                                              0%,,,,,,,,,,,25%,,,,,,,,,,50%,,,,,,,,,,,75%,,,,,,,,,100%
   QUESTION  5              79             95   ************************************************
   QUESTION  6              66             79   ****************************************
   QUESTION  7              54             65   *********************************
   QUESTION  8              44             53   ***************************
```

Figure V.1. A hard copy report of student performance on a cluster of four items.

COMPUTER AIDED INSTRUCTION STUDY MANAGEMENT SYSTEM
DEPARTMENTS OF EDUCATIONAL PSYCHOLOGY AND ECONOMICS
UNIVERSITY OF ILLINOIS
13 FEB 74

STUDENT
PERFORMANCE
ANALYSIS
PAGE 2

AAR
28 CLUSTERS COMPLETED

CLUSTER NUMBER	NUMBER ITEMS CORRECT	PERCENT CORRECT	CLUSTER NUMBER	NUMBER ITEMS CORRECT	PERCENT CORRECT
1	3	75	2	2	50
3	4	100	4	4	100
5	4	57	6	4	100
7	2	66	8	6	100
9	2	66	10	1	25
11	3	60	12	4	100
13	3	60	14	3	60
15	1	25	16	4	100
17	5	83	18	2	50
19	2	40	20	5	100
21	4	80	22	2	50
23	3	75	24	4	100
25	2	66	26	1	25
27	3	100	28	4	80

AVERAGE NUMBER ITEMS CORRECT= 3.1
AVERAGE PERCENT CORRECT= [illegible] 2

Figure V.2. A hard copy report of a student's progress in CAISMS.

The item analysis consisted of a number of indices: 1) the number of students who had completed the assignment to date, 2) the average difficulty and level of interest of the assignment as rated by the students, and 3) an objective measure of item difficulty represented by the mean number of items correct and proportion correct, with standard deviations. Student performance was further analyzed for each assignment to show what number and what percentage of the students got 0 items correct, 1 item correct, 2 items correct, etc. and to show what number and percentage of the students got the first item correct, the second item correct, the third, etc. for each item in that assignment.

At the end of the first semester implementation of CAISMS, the item analysis data were used to help make decisions about procedural changes. For example, difficult items were identified and reviewed to see if they contained ambiguous instructions or information. A number of changes were made with items on the basis of this search. Student performance data (primarily percentages of items correct) were used to make tentative judgements about the project's effectiveness.

We found that the students' ratings of assignment difficulty did not relate well to the objective measures of difficulty. In general, most students, after a short initial period, judged almost all assignments as "average" in both difficulty and interest. It was obvious that students were no longer taking time enough to make thoughtful ratings. We also discovered that we were not collecting some data that would have been useful in further understanding the nature of student study strategies. Primarily, we had stored no performance data from students who had

been required to retake the quizzes. To correct this situation in the second semester dichotomous item data were stored and collected on the initial quiz <u>and</u> the first two retests (when they occurred).

<u>Second Semester Procedures</u>

The log book is still updated daily, and a clerk collects dichotomous item data weekly. Rather than keypunching the data onto cards (which often requires <u>re</u>keypunching many cards) the data are stored directly onto disk via a teletype time sharing system on the IBM 360. This requires keypunching only <u>new</u> data from each student. The analysis programs are internally stored as well, so the time between collection of data by the clerk and running of the analysis varies between 30 and 50 minutes, depending on the amount of student activity during the week.

In the weekly analysis two copies of both student performance data and item analysis data are obtained. One copy is distributed to the course instructors, who have used student performance data for their own information (e.g., who is up to date on their assignments) and for feedback to the students (returning printed records to each student so they can get an overall view of what and how they have done). Student data will again be used in the analysis of student achievement at the end of the semester as a part of the comparative study. Item data will continue to be used for spotting faulty and/or difficult items for review and possible alteration.

Figures V.1 and V.2 show the current form of the analysis for a particular student and a particular assignment. Each student's data appears on a separate page, including his identification, number of

assignments completed, number and percentage of items answered correctly on each assignment, and the average number and average percentage correct for all assignments. For example, Figure V.2 shows that the student named AAR has done 28 assignments, and also shows his performance on them.

An output for an assignment shows the number of students who have completed it, the mean number (and proportion) of questions correctly answered, the standard deviation of questions correctly answered, the number and percentage of students correctly answering 0 items correct, 1 item correct, 2 items correct, etc. with histograms of the indices, and the number and percentage of students correctly answering the first question, second question, third question, etc. with histograms of those percentages also (see Figure V.1).

Two needs exist at present. A more efficient interface between PLATO and the IBM 360 would eliminate the 3-7 hour data collection and re-entry process. Currently we are looking into the possibility of connecting a teletype directly to a PLATO terminal, which will automatically reproduce the PLATO data onto punched paper tape. The data can then be easily entered onto the IBM 360 via the same teletype. The 3 to 7 hour collection period should be reduced to, at most, an hour.

The second need is for on line data analysis, primarily of student data, for use in tutoring, counseling and progress determination (e.g., to determine if a student is ready to take a test). Individual student summary information should be available on PLATO so that up-to-the-minute information can be retrieved. Currently, we are making progress on programming this feature.

VI. SOME REFLECTIONS ON THE USE OF CAISMS

Bruce R. Dalgaard, Economics Instructor

Anticipation Preliminary to Beginning of Instruction

After months of preparation, the beginning of instruction in the CAISMS section of Economics 108 evoked feelings of excitement and a near equal amount of anxiety. Much forethought had gone into the organization of the course material to articulate it with CAISMS. Nonetheless, there was the realization that a lengthy course outline accompanied by page after page describing the functioning of the PLATO system might cause students some anxiety. In addition to this there were the problems involving placement of terminals in the library which necessitated a break in the normal "stay at home" study routine of many students. Also, the initial explanation of CAISMS requirements seemed as though it might convince some students that this much work in the 'dismal science' was more than they bargained for when signing up for a three hour course.

Having a firm belief in the value of the CAISMS project and feeling that total student time outside of class would even out in the long run compared to normal economics 108 classes, it was my desire to convey this to the students. Realizing the antipathy of many students toward new courses and their general distrust of experimental projects, my task was to convey concern for their performance in the course and sincerity regarding the positive aspects of the system.

Equipped with this enthusiastic attitude toward the role of CAISMS in economic instruction and the realization that there were sure to be

problems when a complicated new system is first introduced, the start of the semester aroused mixed emotions.

Initial Student Reaction

My worst fears were realized at the first class meeting. Student enrollment was high, over 70, and members of the class were unaware that computer assisted instruction was to augment the usual lecture-discussion course. Initial student response was negative. The most opposition centered around (in approximate order of intensity) (1) the excessive work load (in terms of hours needed to complete CAISMS quizzes), (2) the utilization of computers and their value as educational tools, (3) the placement of terminals, (4) the requirement to complete CAISMS assignments before being allowed to take examinations.

Given the rather vocal opposition at the first class meeting, it was amazing that more students did not immediately drop the PLATO section of economics. In fact, only five or six students dropped immediately. As a result of a thorough discussion of the CAISMS project, the goals of the course, the role of the computer, and the role of the students, most students were willing to give the project a fair chance.

There were numerous breakdcwns during the first week because of technical problems with PLATO. Students were not properly placed on the system; the system malfunctioned; errors were made in the keying of responses; and students did not use the system properly. Class time was filled with questions and complaints. Student enrollment dropped by another five or six. But, given that there was a high demand for the introductory economics course, enrollment maintained itself at or near

the room maximum of 67. Many of the students who entered the course during the latter part of the first or early part of the second week escaped the technical problems. By the end of the second week the class had settled into a routine and technical problems with the system diminished

Progress of the Students and CAISMS

After the early chaos, the class assumed a normal pace. Because the reading assignments were unevenly apportioned, students had an excessive work load during the first one third of the semester. This, more than anything else, was mentioned by students who dropped the course after the second week, as the major problem. Students progressed at various paces through the CAISMS material (see description elsewhere in this report). Despite occasional malfunctions or human errors, by the first exam all but three students had satisfactorily completed the required CAISMS work, and by the second exam all students had met the requirements.

Student attitudes toward the course ranged from continued belligerence to ecstatic support. Based on a university (rather than project) questionnaire, 61% of the students either agreed or strongly agreed with the statement, "I would take another course that was taught this way." And, 53% disagreed with the statement, "I learn more when other teaching methods are used." When given the opportunity to anonymously make general comments and suggestions, a significant number of students urged improvements in the CAISMS system while approximately 20% suggested that CAISMS be eliminated. Student performance on exams was highly satisfactory. But, since no control group was utilized comparisons are of dubious value. Students often used the system to review rather than for initial learning

by reading lengthy assignments prior to working on the system. Students also found the system to be helpful as a review tool prior to exams. Lengthy review quizzes were placed on the computer and made available to students when they had completed all regular assignments. Students sometimes reported that they did not utilize the system in the manner intended.

Personal Reflections

The PLATO system seemed to be most helpful to students struggling with economics. Occasionally, poor item clusters slowed their progress, but overall, they were able to make use of the system as a learning tool. The students who had little conceptual problem with economics found the system overly time consuming and less than rewarding. Since CAISMS did not employ the language of the text, some students found the questions difficult to answer. These were primarily students who had trouble with the text from the beginning.

After experiencing both the good and the bad elements associated with CAISMS, my overall response is still one of enthusiasm. Further refinements need to be made but, as a means toward improving economic education, CAISMS has great promise.

VII. AN ECONOMIST'S PERSPECTIVE ON CAISMS

Donald W. Paden

The use of CAISMS in the teaching of economics should be viewed against a background of experimentation which is an ongoing commitment on the part of the profession (see Bach, 1973). The typical university, and indeed many colleges, teach economics to hundreds of students each year. The need for innovation is perceived to be twofold: first, to increase student interest in and satisfaction from the first course and, second, to increase the efficiency of the teaching process. In an economy which extols the virtues of rising productivity, the economist is a member of an industry which is a notable laggard in this respect. Thus, anything which facilitates learning for the students or which reduces instructional chores for the instructor is likely to be favorably received by teachers of economics.

During the past decade the need for innovation in the teaching of the first course in economics has led to widespread experimentation both with new teaching methods and with different course content--associated with a variety of delivery systems. With respect to course content there is an almost universal trend toward the exclusion of much of the analytical paraphenalia found in the standard economics textbook. An accompaniment to this trend has been the inclusion of more material dealing with contemporary economic issues than has been true in the past. Thus, relatively few concepts are included in the course and these are illustrated and reiterated in the context of relevant social problems--

poverty, population, the military-industrial complex, communication, consumerism, the multi-national corporation, transportation, urban problems, underdeveloped countries and so on. The core concept idea is part of the teaching strategy at many institutions [e.g., Indiana University (see Saunders, 1973), the University of Oklahoma (Sharp, Leftwich, & Bumpass, 1973), the University of Delaware (Craig & O'Neill, 1973), Stanford University, Carnegie-Mellon University (Journal of Economic Education, 1971, pp. 167-173), Vanderbilt University, and undoubtedly many more]. In a number of schools these courses have been fairly rigorously evaluated against more conventional presentations with favorable results.

Needless to say, different schools have tackled the first course in different ways. For example, at Vanderbilt University, innovation supported by the Joint Council of Economic Education has resulted in the development of a series of case studies progressing from the relatively simple to complex applications of selected core concepts (see Fels & Uhler, 1974). Indeed, the case approach is now available as an option in one of the traditional textbooks (Bach, 1974). Other schools illustrate core concepts by studying relevant social issues from paperbacks containing collections of readings--or, alternatively, from several paperbacks, each covering a specific issue.

In addition to experimentation with limited subject matter, there has been a good deal of activity centered around various delivery systems [e.g., television (see Paden & Moyer, 1972), programmed instruction (Attiyeh, Bach, & Lumsden, 1969; Paden & Moyer, 1972; Weidenaar, 1972), computers (Soper, 1973), games and simulation (Journal of Economic

Education, 1970, pp. 91-129), role playing, etc.]. The two most thoroughly tested of these are television and programmed instruction. In much of the research on television, it is reported that students seem to learn as much as by conventional instruction but like it less. A byproduct of this research raises questions about the contribution of lectures (by television) to student learning (Paden & Moyer, 1972). Independent research investigating live lectures versus no lectures suggests that live lectures do not contribute significantly to student learning at that institution (McConnell & Lamphear, 1969).

The most elaborate research in economics dealing with conventional programmed instruction shows a saving of student time, without significant gains in student learning (Attiyeh & Lumsden, 1969). The use of conventional programmed material as a vehicle for self-instruction continues, however, in many schools with satisfactory results (e.g., Sharp, Leftwich, & Bumpass, 1973).

During the past several years, the University of Oregon has accumulated evidence of the effectiveness of the interview method and the use of undergraduate tutors (see Newton & Abraham, 1973). An even sharper break with tradition has been supported by the Joint Council on Economic Education at the University of Colorado (see Boulding & Boulding, 1973). Here Kenneth and Elise Boulding have devised a course titled "An Introduction to the Economic Systems." The course deals not only with economics but, history, demography, technology, communication, resources, social welfare and, among other things, religion. The attempt is to orient the student in the total system imposed upon the individual by these environmental facts of life.

As seen from the standpoint of the user (in this case the Department of Economics at the University of Illinois) CAISMS I and, hopefully CAISMS II promises to evolve into a way of tailoring the first course in economics to fit the requirements of large numbers of students with a wide range of ability, motivation, study habits, interests and background. Given the availability of a computerized test-bank, students with sufficient background can proficiency out of the principles course the first week of class and can be scheduled into advanced courses. Within limits, students will be able to complete the course whenever they achieve a prescribed level of proficiency. Average or poor students can be placed in learning situations appropriate to their ability. Better students can be channeled into seminars studying topics which are of special interest to them. Moreover, there seems to be no reason why television, games, simulations, campus lectures, etc. cannot be integrated into the course without undue management problems. Thus, it may be possible to treat students in large classes as individuals, taking advantage of the printed message, personal contact, peer interaction and a variety of technological innovations (see Kelley, 1973).

References

Attiyeh, R., Bach, G. L., & Lumsden, K. G. The efficiency of programmed learning in teaching economics: The results of a nationwide experiment. _American Economic Review_, 1969, _59_ (2).

Bach, G. L. An agenda for improving the teaching of economics. _American Economic Review_, Papers and Proceedings, 1973, _68_, 303-308.

Bach, G. L. _Economics_. Princeton, N. J.: Prentice-Hall, 1974.

Boulding, K., & Boulding, E. Introducing freshmen to the social system. Unpublished manuscript, University of Colorado, 1973.

Craig, E. D., & O'Neill, J. B. Providing choice overcomes the dismal science syndrome. Unpublished manuscript, University of Delaware, 1973.

Fels, R., & Uhler, R. G. _Casebook of economic problems and policies_. St. Paul, Minn.: West Publishing Co., 1974.

Journal of Economic Education, The. Articles on games and simulation, 1970, _1_, 91-129.

Journal of Economic Education, The. Special report: An experimental evaluated elementary course, 1971, _3_, 167-173.

Kelley, A. Individualized instruction through the use of technology in higher education. _The Journal of Economic Education_, 1973, _4_, 77-89.

McConnell, C. R., & Lamphear, C. Teaching principles of economics without lectures. _The Journal of Economic Education_, 1969, _1_, 33-45.

Newton, J. N., & Abraham, F. J. An exploratory study of the effectiveness of alternative instructional procedures in the teaching of economics. Unpublished manuscript, University of Oregon, 1973.

Paden, D. W., & Moyer, M. E. The teaching of the first course in economics. Research Papers in Economic Education, Joint Council on Economic Education, 1972, 85-92.

Saunders, P. Experimental course development. Unpublished manuscript, Indiana University, 1973.

Sharp, A. M., Leftwich, R. H., & Bumpass, D. L. Experiments with basic economics. Unpublished manuscript, University of Oklahoma, 1973.

Soper, J. C. Programmed instruction in large-lecture courses. The Journal of Economic Education, 1973, 4, 125-129.

Weidenaar, D. The effect of programmed instructional materials and other factors on performance in the introductory economics course. Research Papers in Economic Education, Joint Council on Economic Education, 1972, 118-132.

VIII. COST PROJECTIONS ON THE DEVELOPMENT AND USE OF CAISMS

Thomas H. Anderson

In our early conceptions of CAISMS, we contended that CAISMS would be relatively inexpensive to develop and operate. One rather naive way to index cost would be to simply divide the dollars in the grant by the instructional output. Since we developed about ten instructional hours of terminal time per student with a $100,000 grant, the index computes to $10,000 per hour of instruction.

It must be remembered, however, that other tasks were also accomplished in that grant year which are only indirectly related to the production of instruction time. These tasks include, 1) laboratory and literature research related to questioning procedures, 2) the development of an evaluation plan to assess the effectiveness of CAISMS, and 3) the development of a study management system which can be used in a variety of course settings. Presented in the remainder of this section are the most realistic time on task and cost data that we can compute concerning 1) the development of items, 2) the program planning and coding of Tutor, and 3) the costs and maintenance of hardware.

Time Estimates to Develop Items

A team of economists and educational psychologists developed a pool of approximately 300 questions. A detailed discussion of the team's operation is presented elsewhere in this report. Since eight persons were involved at different times with the item development project, it's very difficult to obtain highly accurate indices. We think the estimates are in the ball park, however.

1.	Subject matter experts	200 hrs.
2.	Educational specialists	250 hrs.
3.	Students who tried out items	200 hrs.
	Total hours to develop 300 items	650

Time Estimates for Program Planning and TUTOR Coding

A team of three educational psychologists and one undergraduate clerk developed this phase of the project. None of the three persons in the programming team was on the item development team. It should also be mentioned that none of the programming team knew the TUTOR programming language at the onset of the project.

1.	Off-line development	100 hrs.
2.	On-line development	225 hrs.
3.	Maintenance and monitoring activities	100 hrs.
4.	Cleric file building	25 hrs.
	Total hours devoted to program development and coding	450

The Costs and Maintenance of Hardware

The latest hardware and maintenance costs are:

1.	Purchase price of a PLATO terminal	$5,335
2.	Communication hookup	288
3.	Phone line	100 per year
4.	CPU usage	1,500 per year
5.	Maintenance	50 per year

In an effort to compute a comparative index, the terminal cost was depreciated over a five year period and all costs were estimated on a per year basis. The resulting estimate is $2,750 per year per terminal. Using our rather generous estimate of eight terminals per 100 students, an index of $77 per semester per student is the resultant. A more realistic estimate of CAISMS terminal usage would be about 200 students per eight terminals. The dollar index would be reduced to $38 per semester per student in this case.

Course Development Cost Estimates

Presented below are three course development cost estimates.

First semester costs of developing and implementing CAISMS for 100 students in Economics 108 are:

1)	Development and maintenance of system (1,050 hrs. @ $10/hr.).....................	$10,050
2)	Hardware and maintenance ($38.50/student)..	3,850
		$13,900 per 100 students
		or $140 per student

Second semester costs of implementing the same economics course for 100 students are:

1)	Systems maintenance (40 hrs. @ $10/hr.)...	$ 400
2)	Hardware and maintenance ($38.50/student).	3,850
		$ 4,250 per 100 students
		or $42.50 per student

Costs of developing and implementing each additional new course are:

1)	Item development (600 hrs. @ $10/hr.).....	$ 6,000*
2)	Data entry specialist (25 hrs. @ $2.50/hr.)	65
3)	Programming system maintenance (40 hrs. @ $10/hr.)........................	400
4)	Hardware and maintenance ($38/student)....	3,850
		$10,315 per 100 students
		or $103 per student

---

*This cost may be much lower if items are available and do not need to be developed from 'scratch'.

The Comparative Costs of CAISMS and Other Methods of Instruction

It should be emphasized that the conventional small group (maximum N=30) lecture-discussion technique is an expensive method of instruction. A cost figure per student assuming classes of 30 students and one-third of the instructor's salary allocated as instructional cost is $155. This estimate is based on a $14,000 nine month salary of an assistant professor. These estimates are computed from a scheme used by Paden (1962).

The fact that small group instruction is so expensive makes it necessary for many institutions to use large group lecture sessions. Assuming a salary of $21,000 for a full professor's nine month salary and one-half of that salary is applied to lecture responsibilities, the cost for 300 students is approximately $35 per student. The large group lecture method is subject to decreasing cost as the number of students increases (a minimum of $23 per student) until the lecture group becomes so large that a second lecture becomes necessary. Costs for the second group are comparable to the first group. There are numerous qualifications which need to be made. For example, the cost of paper grading, test scoring, student conferences and so forth has not been included in the estimates.

The costs of the small group and the large lecture section techniques of instruction are presented in Figure VIII.1. In addition, the additive costs of using CAISMS as a supplement to the large lecture technique are plotted relative to the two conventional methods. In general, to use CAISMS doubles the cost of large lecture instruction, but the cost of a combination of large lecture instruction and CAISMS is approximately one-half that of small group instruction.

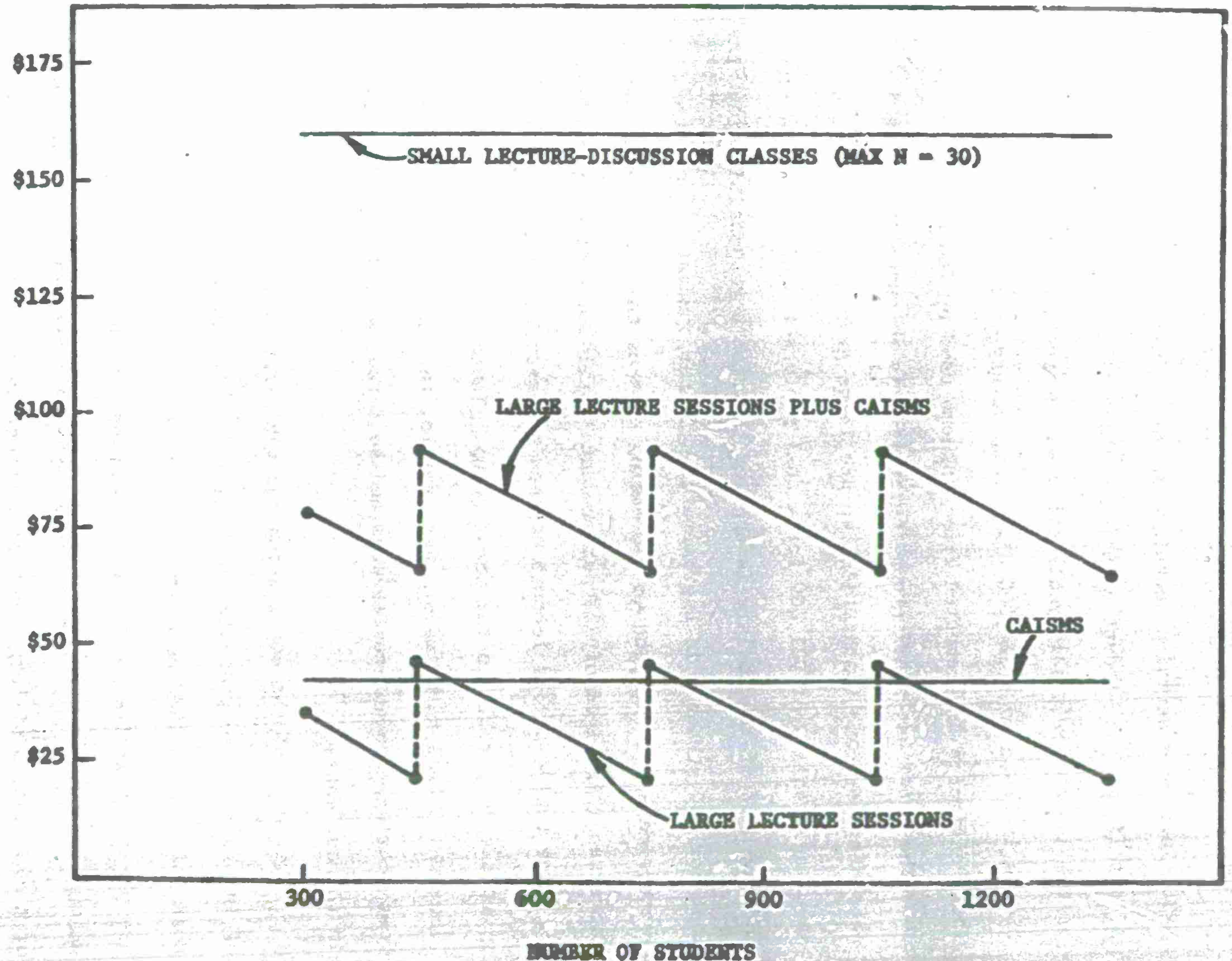


Figure VIII.1. Cost per student for three credit hours of instruction.

References

Paden, D. W. The teaching of economics via television at the college level. Supplement to J. W. Meaney, Televised college courses--A report about the college faculty released time program for television instruction. New York: Fund for the Advancement of Education, 1962.

IX. CAISMS II: A PROPOSED INNOVATIVE COURSE MANAGEMENT SYSTEM

Richard C. Anderson, Donald W. Paden, Bruce R. Dalgaard,
and H. Richard Smock

Currently being planned is an extended version of CAISMS which we hope will significantly alter the character of post secondary education. Although our special focus in this proposal is on the large introductory university course, we are confident somewhat similar plans could be made to work in community college, military, or industrial settings.

Here is a sketch of the system we envision. Conventional lectures and standard quiz sections will be scrapped. These are typically dull, ineffective, and inefficient. They squander human resources. Instead, the student will be expected to acquire the basic information and concepts entirely from reading. His attention to the material will be maintained and his progress monitored by the already-functioning study management system.

The time of instructors, saved because lecturing and routine review are no longer required, will be deployed in remediation for students having trouble mastering the core curriculum and a smorgasboard of topical seminars and special projects for the rest. These activities are expected to provide depth not ordinarily reached in an introductory course; to make real and concrete what would otherwise be merely abstract and verbal; to spark the interest of most and the excitement of a few.

CAISMS II will break the tyranny of the university calendar. It will be possible for a student to complete a lean course in as little as eight or ten weeks for three hours credit. Or, he can complete the enriched program, working harder and longer, earning perhaps five credit hours.

The program will feature a mastery learning strategy (provided Department approval can be obtained). Virtually no student who is willing to continue trying will get a grade lower than B. If a student falls below B-level mastery on his first attempt at an exam he will be given extra help and a second chance to take the exam, or even a third or fourth chance if he needs it and has the motivation to work on.

CAISMS II will accommodate a wide range of student individual differences in interest, ability, and energy. The slow student gets a forgiving system, one that does not punish him excessively for failing to reach criterion easily. The able student gets options. He may complete the course quickly. Or, he may accept the opportunity to explore the field in more depth.

A free-access examination system is required in order for the scheme which has been outlined to function properly. Within a few, broad limits students will be able to take graded examinations whenever they are ready and willing. Otherwise, the program would lose its flexibility. We have conceived a computer-based plan which would allow an eligible student to take an exam any time during the week, which would be relatively cheat proof, and which would make only modest demands on instructors for grading. This plan will be detailed in the next section.

CAISMS II is expected to have several benefits. The first is an increase in achievement. So far as performance on course examinations is concerned, striking improvements are expected in the performance of students at the low end of the ability distribution. Many able students are expected to learn a good deal more than usual as well, although because of

the limitations of educational measurement these gains may not be as readily apparent.

The second benefit should be heightened student interest. Students will be receiving instruction closely tailored to their individual requirements and interests. Instead of a lock-step system, they will get choices. Able students will be involved in seminars that are relevant and intellectually challenging. Instead of hearing the professor from the rear of the lecture hall, they will have chances to talk to him in person about topics of mutual interest, and to explore topics which allow application of knowledge acquired through reading. It stands to reason that most students will appreciate these features. A higher than usual incidence of genuine enthusiasm seems likely. Such evidence as is now available supports this hope, as will be explained in the next section.

Third, in comparison to other individualized, mastery learning systems, CAISMS II will require much, much less clerical labor from instructors. This is important in the first place because it means more intelligent use of trained personnel. Even more important in the long run, if an innovation is to be accepted, it must provide satisfying, productive roles for people. Finally, we guess that, discounting development costs and equipment costs, an institution could operate CAISMS II at about the same price as a traditional instructional program.

Components of CAISMS II

This section will describe the key elements of CAISMS II in more detail. Several of the components have been widely tried elsewhere and thus there is evidence regarding effectiveness and attractiveness to students.

This evidence will be briefly reviewed. More important, endemic problems have become apparent. These will be discussed. We shall attempt to show that CAISMS II can ameliorate these problems. Specifically, this section will consider a) whether lectures are dispensible, b) the probable impact of a mastery learning scheme, c) difficulties with individualized instructional systems, d) the rationale for a computer-based, free-access examination system, 3) remedial instruction, and f) the problems and prospects for seminars.

Are Lectures Dispensible?

The answer to this question is an emphatic yes. The June 1972 issue of the Personalized System of Instruction Newsletter reported a questionnaire study of 190 courses in which students were expected to learn from textbooks and other self-instructional activities. In most of the reported cases lectures had been eliminated entirely. The disciplines most heavily represented were psychology, physics, engineering, mathematics, chemistry, and biology. There was a smattering of courses from such other fields as Spanish, English, and sociology. Almost all respondents claimed students learned more from "personalized" (i.e. individualized) instruction. It was reported that students heavily favored the self-instructional to the lecture mode. Only 3 of 190 courses were said to have been failures on balance.

Experimental comparisons have also consistently shown higher student achievement under individualized instruction than under traditional lecture or lecture-discussion methods, according to a recent, comprehensive review of the literature by Parsons (undated). Moreover the experimental

literature indicates that individualized, nonlecture courses are very popular with students, even though usually rated as more work than conventional courses.

Two of a rather large number of comparative studies will be detailed for illustration. Witters and Kent (1972) compared traditional and individualized introductory psychology and cultural anthropology courses. Students in the experimental section of the psychology course had no lectures, indeed, no formal class meetings of any kind. Students participating in the individualized anthropology course met as a group once a week to view movies. Results showed significantly higher achievement, especially for students with low cumulative grade point averages, in the nonlecture sections of both courses in each of two semesters. Students in experimental sections consistently reported greater confidence in their mastery of course material. In both semesters of the psychology course and in one semester of the anthropology course, students whose instruction was personalized rated the course as significantly more enjoyable than their counterparts in the lecture-discussion sections. Paden and Moyer (1972) compared several different methods for teaching the first course in economics. Most noteworthy was the fact that students who received the standard course minus the lectures achieved as much as those required to attend lectures.

There is a lot of evidence that students don't go to lectures unless they have to. For instance Lloyd _et al_. (1972) studied the "reinforcing" properties of university lectures. They studied frequency of attendance as a function of whether 1) quizzes were given, 2) material to be covered

in quizzes was discussed, 3) the lectures were on reading assignments, or material unrelated to quizzes, 4) credit was given for lecture attendance, 5) information on future quiz questions was provided or 6) attendance was made contingent upon completion of reading assignments. On days when quizzes were given or when impending quizzes were discussed, student attendance was above 90%. Attendance fell as low as 5% when the lectures were over reading assignments or unrelated material. When students were then given grade-related points for attending lectures, attendance again jumped to above 90%. Under the condition in which students were eligible to attend lectures only if caught up on reading assignments, attendance averaged only 25% of those eligible.

In summary, even a cautious person, making allowance for the somewhat evangelical fervor of advocates of the Personalized System of Instruction, would conclude that a college or university course can be effective and attractive without lectures. Needless to say, courses which depend upon students learning primarily from reading are not without problems. These will be considered after the section on mastery learning since in virtually all of the recent experience the nonlecture mode and mastery learning have gone hand in hand, at least at the postsecondary level.

Probable Impact of a Mastery Learning Strategy

The mastery learning scheme is very simple. A standard, or mastery criterion, is set for each unit of work. The student who achieves the criterion moves on to the next unit. The student who does not receives further instruction and then takes the examination again. There is no penalty for failing the unit test, but the student is not permitted to go on to a new unit until the current one has been mastered.

Mastery learning is not a new idea. It was the heart of the famous Winnetka Plan developed fifty years ago by Carleton Washburne (1922). In recent years Bloom (1968, 1973) and his students (e.g. Block, 1971) have done the most research and writing on mastery learning.

According to Block (1971) the Winnetka Plan was popular during the 1920's and 1930's but eventually faded because of technological problems. The constant grading of examinations was an irksome chore for teachers. Students spread out over so much of the course-of-study that it was very difficult to deal with them all in a single classroom. A mastery learning strategy is not feasible unless instruction can be individualized.

Regarding the effectiveness of mastery learning, Block (1971) in his book, Mastery learning: Theory and practice, concludes as follows:

> The results from approximately 40 major studies carried out under actual school conditions ... [indicate that] ... in general, three-fourths of the students learning under mastery conditions have achieved the same high standards as the top one-fourth learning under conventional group based instructional conditions. In studies where a strategy has been refined and replicated, 90 per cent of the mastery learning students have achieved as well as the top 20 per cent of the non-mastery learning students. Mastery learning students also have exhibited markedly greater interest in and attitudes toward the subject learned compared to non-mastery learning students (pp. 8-9).

One specific study will be described for illustration. Kim et al. (cited in Block, 1971, pp. 124-126) compared mastery and nonmastery strategies in an experiment that entailed eight weeks of instruction in English and mathematics for approximately 5,800 seventh graders in nine schools in Seoul, Korea. Programed self-instructional booklets were employed to permit individualization. The pupils graded each others unit quizzes. The bright pupils were called upon to tutor the slow ones. While there

was considerable variability among schools, the overall results indicated that a much higher percentage of the mastery than the nonmastery pupils reached the criterion of 80 per cent on the final examinations. In English 72% of the experimental subjects reached criterion as compared to only 28% of those learning under ordinary instructional conditions. In mathematics, an average of 61% of the mastery as compared to 39% of the nonmastery students attained the final test criterion. It is reported that a study involving 32,000 students is underway.

As already noted, the Personalized System of Instruction also requires the student to master one unit before he is allowed to undertake the next; however, personalized instruction has had virtually no contact with the mastery learning tradition discussed in this section. The truly extraordinary insularity of educators and psychologists (maybe Skinnerians especially) is evidenced by the fact that only 2 of the 49 items appearing in Block's (1971) annotated bibliography are listed among Parson's (undated) 72 references!

In summary, a strong case can be made for a system of mastery learning. It should be emphasized that the current version of CAISMS already involves mastery learning on a microscale. The student must pass an ungraded quiz over each short reading assignment before he is given the next assignment. CAISMS II will additionally involve mastery learning on a macroscale. That is, the student will have to pass graded examinations over large units of course material before proceeding.

Difficulties with Individualized Instruction System

This section will consider problems with individualized instructional systems and the steps we propose to solve them in CAISMS II.

High drop out and failure rates. While much larger proportions of students achieve high levels of competence under the kind of instructional regimen we are discussing, there is also an increase in proportion of failures. Instead of the usual bell shaped function the grade distribution is shaped like a "U". Few people get C's. Most get A's. A somewhat higher than usual proportion flunk (Parsons, undated, p. 17). Furthermore, the dropout rate in individualized college courses run on an individualized, mastery learning basis is three to five times as high as it is in conventionally taught courses (Parsons, undated, p. 19).

Why failure and drop out rates should be high is easy to understand. In the type of course being considered the student *must* complete his work and actually master it. He cannot get away with walking into an examination, crossing his fingers and hoping for the best. The lazy, poorly motivated, or disorganized student either gets himself in gear, or he fails. It's as simple as that.

Born and Whelan (1973) found that while most students who dropped an individualized introductory psychology course had low cumulative GPAs, 80% of them had passed the most recent unit quiz taken. So, they were not actually failing. The most notable characteristic of the students who withdrew was that without exception they were behind the work schedule needed to complete the course by the end of the semester.

It will come as no surprise to an experienced college teacher to learn that research shows that very large numbers of college students do very little studying and then only under the immediate threat of a graded examination or project. In a study we have completed, and will eventually

issue as a technical report, 55% of a sample of 202 University of Illinois undergraduates indicated that they study only when circumstances (tests, assignments, etc.) require. Another 14% reported studying rarely, if ever, or in "a most irregular manner." Studies by Brown (1964) and Brown and Dubois (1964) indicate that there are moderate to high correlations between work habits and study methods and grade point average across a range of science and humanities courses. The tendency to avoid delay and get started with work as soon as it is assigned is an especially good predictor of achievement.

Probably one reason individualized mastery learning systems are successful is that a higher proportion of students get down to business earlier and work more systematically. But there is still a group not reached. We are attempting to design CAISMS II to reduce the drop out and failure rates associated with similar systems.

Criterion of mastery. Instead of setting the mastery criterion at 100% or even A-level performance, which is the common practice, we shall be satisfied if everyone does at least B-level work. This will keep slower students from endlessly repeating units.

Incentives and sanctions. Incentives will be made available for completing work on a reasonable schedule and sanctions will be enforced to prevent students from falling behind some minimum schedule. Several incentives are being considered.

First, those who master units of work by established cut off dates will become eligible for seminars and special projects. These will be geared to student interests and thus should be an attractive incentive to many.

Second, such students will be given the option of taking the course for additional credit.

Third, under the assumption that students who maintain a reasonable work rate have demonstrated self-discipline, the external discipline imposed by the study management system may be relaxed somewhat. The system is designed so that at least seven minutes, the time judged necessary to thoroughly study the typical reading assignment, must lapse from the time the student receives an assignment until the system will give him the quiz. This procedure forces the student to study near a terminal, to read and then to answer questions immediately, which is exactly what we want; the attention-maintaining and review effects of questions are maximized when the student is under their proximate control.

Deep processing of text could probably be maintained with some lesser degree of control for disciplined students. If the programed delay between assignment and quiz were cut to, say, three minutes for such students (but not those lagging behind) an alternate study strategy would become possible. A student could read a large block of material in his room or at home on a weekend, sign in later at a PLATO terminal, receive an assignment (which he had already read) and three minutes later take the quiz. The three minute delay would encourage review and prevent abuses of the system.

As for sanctions, we expect to use a "Doom's Day contingency" (Malott & Svinicki, 1969). The student who has not taken a graded examination by an irrevocable but lenient cut-off date will automatically receive an E for the course. This is not a cruel or unusual requirement. An unexcused absence from an examination is grounds for failure in any college course.

The Doom's Day contingency is simply an extension of this policy to an individualized instructional system. The catch for the lazy student in CAISMS is that he must make himself eligible for graded examinations by completing required work. There is no way he can go into an examination unprepared, praying for good luck.

In many individualized college courses the student is asked only to have completed the required work by the end of the semester. CAISMS II will enforce at least a minimally satisfactory rate of work. Graded examinations to which the Doom's Day contingency applies will be required periodically throughout the semester.

It is painfully clear to observers of the college scene that large numbers of students, including many who are able and reasonable conscientious, do not really get down to work until they must (cf. Anderson & Faust, 1973, pp. 254-260). Many individualized instructional systems have not forstalled procrastination. The great latitude permitted students may even encourage putting study off until the end of the semester. Barr and Jakobson (1973, p. 5) described the problem in graphic language when explaining to new students in an individualized physiology course a rule which required work to be completed earlier: "If you had been here the last week of the first semester ..., you wouldn't have to ask why we have established this requirement. The only longer lines in the country were at the World Theatre in New York, where Deep Throat was playing at the time."

Naturally we would not care to claim that the Doom's Day contingency itself will prevent drop outs and failures. In fact, it will surely be a cause of grief for a few students. The net effect should be favorable,

however, Procrastination will be reduced, and procrastination appears to be a principal cause of eventual poor performance under individualized, student-paced instruction.

Counseling for laggards. A third procedure to reduce the number of withdrawals and failures is special counseling for students who fall behind the slowest acceptable work rate. This past fall a graduate student in counseling psychology, in collaboration with the instructor in the experimental section of Economics 108 and other members of the CAISMS staff, worked with a group of students who were not making satisfactory progress. The hope was to develop a program to help the procrastinating student "get it all together," clarify his goals, acquire more systematic methods of study and techniques for self-management. Based on self-reports of the students involved and objective measures of subsequent performance, more than half of the "laggards" was helped. The program will be continued. The idea is for the teacher and his associates to take initiative and seek out the lagging student before he gets into serious trouble.

Clerical and logistical problems. Individualized, mastery learning programs entail a heavy burden of clerical work. There must be means to give and score quizzes during many hours of the week. Fairly elaborate records of student progress are required. Tutorial and small group sessions for shifting subsets of the student population must be scheduled.

In many cases the burden of an individualized system has proved too great for faculty and graduate assistants. The typical solution is undergraduate "proctors" or "tutors." They may be students doing well in the course, or students who have taken the course in a previous semester.

They are either paid by the hour or receive course credit. Undergraduate proctors can work out well. Indeed, if there proves to be too heavy a load of remedial instruction for graduate assistants, we will consider using them for this function in CAISMS II.

The computer will do the bulk of the clerical labor in CAISMS II. It will also handle the logistics of scheduling review sessions, seminars, and even individual appointments. The details will be discussed in subsequent sections. In the meantime, it seems reasonable to hope that an interactive computer system can solve the most annoying set of problems (see Parsons, undated, p. 19) that has arisen with respect to individualized college instruction, problems which in the past has caused programs with a similar conception to flounder (Block, 1971, p. 4).

Computer-based, Free Access Examination System

To be developed is a system which will allow an eligible student to take a graded examination whenever he has access to a supervised PLATO terminal. The system will work like this. Any time after the student has made himself eligible by completing required work, he may request the examination. (The current version of CAISMS already keeps track of eligibility for off-line examinations.) Part of every examination will consist of multiple choice and short answer items. These will be administered at the terminal and scored instantly. Examinations may also entail essay questions, which will be administered by the computer but answered on paper and scored by a human teacher.

We insist on the capability for essay examinations, despite such problems as unreliable scoring and demands on teacher time, because we are

convinced that mastery of such subjects as economics and history cannot adequately be assessed solely with "objective" tests. Furthermore, we believe that an undesirable change in the character of student study is not unlikely were he to prepare only for multiple-choice and short-answer tests (cf. Myer, 1934, 1935; McKenzie, 1972). Integration of concepts and principles seems less likely; the student might get by, or think he can get by, with learning information in a piecemeal fashion.

Cheating. Several techniques will be used to minimize cheating. First samples of multiple-choice and short-answer items will be drawn at random from a large pool. This means every student will receive a somewhat different exam. For instance, an exam might consist of 50 items drawn from a pool of 150.

One of the participants in the CAISMS project, Donald Paden, has prepared 500 economics achievement test items, so a good deal of the work to implement an item sampling procedure in the economics course has already been done. Item analyses giving the difficulty level of every item are available. This means that it will be possible to equate the different examinations of the various students by weighting each individual's score according to the difficulty of the items he happened to receive.

To reduce further the likelihood that a student could learn much by looking over another student's shoulder, 1) there will be alternate forms of each item which differ only in the order response alternatives are listed, 2) the items will not be numbered or otherwise identified, and 3) the items will be paced, allowing enough time for anyone to read and think but not so much that the item could be copied.

Gross forms of cheating, such as one person taking an exam for another, will be discouraged by spot checking student ID cards. While we very much hope that we will have a few terminals to install in such places as residence halls and local public libraries, exams will be given only at central installations--in the Undergraduate Library at the University and the Learning Resources Center at Parkland College. This will make spot checking feasible and also bring social pressure to bear. Only the brazen would seek or give help when there were many other people around.

A final point with respect to cheating: under a mastery learning regimen the consequences of failure are less awful. You'll get another chance. There should be less incentive to cheat.

Essay questions. The procedure envisioned to handle essay questions is a modified version of the take-home exam. The student is given a short list of questions when he begins a unit. When he is ready for the exam the computer randomly assigns him one (or more if desired) of these questions. A fixed period of time is allowed to write, whereupon the student turns in his answer to someone at the library reserve desk or to the attendant on duty at the learning resources center, who stamps it with a printout clock. Of course the CAISMS system has a record of when the student received the question. Periodically the essays are collected for grading.

The teaching personnel will specialize in grading certain of the essay questions. This practice will improve scoring reliability and make the task less burdensome. Another device to save teachers' time is to permit a student to attempt the essay portion of an exam only when he has passed the objective portion. When an essay has been read, the grader will enter

the mark and the completion time into the system. CAISMS will check to see whether the student finished the exam within the time limit, and will instigate the appropriate action depending upon whether the student passed or failed.

This is a good place to mention that CAISMS II will require a PLATO terminal in the course office. A number of functions will be served including clerical tasks such as entering grades, instructional management decisions that require convenient access to student records, some important teaching functions that will be detailed in the next section, research functions that entail access to data, and the authoring of course materials. It would be desirable if this terminal were fitted with a device for producing hard copy records.

<u>Alternate examination system</u>. The system described in the preceding section will make heavy peak load demands on terminals. If funds to purchase enough terminals are not found in time, we will temporarily employ an alternate scheme in the Fall of 1974.

For each graded examination, as many samples of items as there are students in the course will be randomly drawn by a computer and outputted by a line printer. The correct answers will be printed in the right margin in such a way that they can be chopped off with a paper cutter. A clerk will be hired to give an exam to any eligible student who appears at the course office during business hours. Eligibility can be checked at the terminal in the course office. The clerk will score the test and enter the score into the system. Essay questions will be assigned by the clerk using a list of random numbers.

Remedial Instruction

Absolutely essential for an individualized, mastery learning system is effective back-up instruction when the student fails on his first attempt. Let us be clear that we are talking in this section about remedial instruction to be provided after poor performance on a graded unit examination. CAISMS II will continue to invoke the simple contingencies already operating when a student does not pass the ungraded quiz over a short reading assignment. In other words CAISMS II will have components to deal with poor performance on both a microscale and a macroscale.

Diagnosis. We believe that we are in a position to be especially successful at providing remediation. The first requirement is good diagnostic information, both about the difficult aspects of the subject matter and about the misconceptions of individual students. With respect to the former, item analyses of the data collected in the initial runs of CAISMS will give a virtual blue print of the difficult sections of the textbook. This information will supplement the informed judgement of experienced teachers. With respect to diagnosis of the difficulties of an individual student, useful diagnostic information will be an almost automatic by-product of the on-line examination system. Test items will be coded according to the concept assessed. A topic by topic profile of the student's performance can be displayed on the PLATO screen. The student and his teacher can see what the problem is. This should be a big help since "a few minutes of instruction which gets precisely at the student's difficulty can be worth more than hours of general review" (Anderson & Faust, 1973, p. 176).

Alternate teaching procedures. A second requirement for successful remedial instruction is effective alternate teaching procedures. These will be prepared in advance for topics which are judged likely to prove difficult for many students. The techniques to be used are individual tutorial, small-group review sessions, fully on-line CAI lessons*, and small step, programed text material. Closely programed material is known to be especially effective with students who have difficulty understanding a point when other forms of instruction are employed. One of the CAISMS group, Donald Paden (1972), has authored a programed introduction to economics.

Parenthetically, when our experience this year indicates that large numbers of students are having difficulty getting a point from the text alone we shall not depend on repairs once an exam has been failed. Rather, we will supplement the text beforehand. Marginal analysis is an apparent example of such a difficulty in Economics 108. Either the concepts are inherently subtle or complicated, or the text is opaque. Techniques for handling such problems will include asking more and better CAISMS questions, augmenting the text with fully on-line CAI lessons, assigning material from other texts including programed texts, or scheduling everyone for lecture-discussion sessions.

Human contact. When a student fails an exam, he is likely to be discouraged, anxious, angry, or even depressed. A computer is a marvelous instrument, but it is not likely to be very comforting. This is one reason

---

*We do not plan to invest heavily in CAI lesson development. Rather we will partially subsidize interested graduate students and faculty.

we are planning a human contact whenever a student fails. The student will get the bad news at a PLATO terminal and then be asked to make an appointment with a tutor. This he will do by entering his name on a calendar displayed on the PLATO screen showing available dates and times.

When the student meets with his tutor, they will go over the problems the student had with the exam, using as an aid the topic by topic analysis of his performance displayed on the terminal in the course office. The tutor will prescribe remedial work of one of the types listed above. If the problem appears to be a small one it could be cleared up at this first meeting. In any event, the student will become eligible to repeat the exam when the tutor is satisfied that the student's difficulties have been surmounted. By "repeating an exam" we mean of course receiving a new random sample of items which will duplicate ones in the first exam only by chance. Of course, students progressing at a satisfactory rate will also be able to meet individually with a course instructor to discuss matters related to the course. Any student will be able to make an appointment with a teacher (at his initiative) from a PLATO terminal.

Seminars: Problems and Prospects

It is perhaps a portent that Fred Keller's (1968) seminal paper on personalized instruction was entitled "Goodbye, teacher..." For now, in a "personalized course," the professor is no longer a lecturer; he is no longer a teacher of any kind. He is called a "manager," and that he is, not only in the Skinnerian sense as the arranger of reinforcement, but in the ordinary sense as well.

CAISMS II will employ teachers as teachers, under conditions which make scholarly insight and human qualities relevant. Seminars will be offered in which class size is kept (relatively) low. Only students who have mastered required background work will attend, so the discussion should be informed.

Many college students, especially at large state universities, yearn for more opportunity for give-and-take. As one student evaluating a special course at the University of Illinois put it,

> I think courses like this are needed because of their informal atmosphere. They really let you hang loose and express yourself. They provide an atmosphere for an interchange of ideas... (Smock & Myrow, 1971).

The program of seminars should provide for interaction among students and between students and teachers, help make concrete abstract aspects of the subject matter, provide real depth in a few topics and, hopefully, increase the enthusiasm of the students toward the course and the subject matter.

Seminars typically will meet one hour a week for three or four weeks. To be eligible to enroll in a given seminar the student will have to have mastered a specified amount of the core curriculum by a certain registration date. The seminar schedule will be staggered so as to accommodate variations in student work rate. Popular seminars will be run several times during the semester, at different times of the day.

CAISMS will handle the scheduling of seminars. As soon as a student becomes eligible the system will inform him of this fact and describe his options. Each seminar will have an enrollment limit. Positions in the seminar will be subscribed on a first come first serve basis. This will be an added incentive for completing work quickly.

At any one time in a course of 300-350 students there might be seven or eight seminars in operation. Typically, however, only half of the students will be enrolled in a seminar at a given moment. Early in the semester slow students will not have become eligible. Late in the semester fast students who are electing the minimum number of credit hours will have completed course requirements. Hence, seminars will range in size from 20 to 30, assuming that four half-time graduate assistants and one professor comprise the course staff.

To give a concrete example, one seminar in Economics 108 might involve the functioning of the commodity futures market. Students would be able to explore a case that approaches perfect competition. A study of the futures market would also highlight the intensifying world wide demand for food products. Several nonconventional teaching techniques could be considered in addition to reading and discussion, such as a market simulation game played on PLATO, or a visit to the Mercantile Exchange or Board of Trade. On such a trip, informal discussions with professionals in the field could be arranged.

References

Anderson, R. C., & Faust, G. W. Educational psychology: The science of instruction and learning. New York: Dodd, Mead & Company, 1973.

Barr, L., & Jakobson, E. Introduction to human physiology. Champaign: Stipes Publishing Company, 1973.

Block, J. H. (Ed.) Mastery learning: Theory and practice. New York: Holt, Rinehart and Winston, Inc., 1971.

Bloom, B. S. Learning from mastery. Evaluation Comment, 1968, 1, No. 2.

Bloom, B. S. Recent developments in mastery learning. Educational Psychologist, 1973, 10, 53-57.

Born, P. G., & Whelan, P. Some descriptive characteristics of student performance in PSI and lecture courses. Psychological Record, 1973, 23, 145-152.

Brown, F. G. Study habits and attitudes, college experience and college success. Personnel and Guidance Journal, 1964, 43, 287-292.

Brown, F. G., & Dubois, J. E. Correlates of academic success for high-ability freshman men. Personnel and Guidance Journal, 1964, 43, 603-607.

Keller, F. S. "Good-bye, Teacher..." Journal of Applied Behavior Analysis, 1968, 1, 79-89.

Lloyd, K. E., Garlington, W. K., Lowry, D., Burgess, H., Evler, H. A., & Knowlton, W. R. A note on some reinforcing properties of university lectures. Journal of Applied Behavior Analysis, 1972, 5, 151-155.

Mallott, R. M., & Svinicki, J. G. Contingency management in an introductory psychology course for one thousand students. Psychological Record, 1969, 19, 545-556.

McKenzie, G. Some effects of frequent quizzes on inferential thinking. American Educational Research Journal, 1972, 9, 231-240.

Myer, G. An experimental study of the old and new types of examinations: Part I, Effects of examination set on memory. Journal of Educational Psychology, 1934, 25, 641-669.

Myer, G. An experimental study of the old and new types of examinations: Part II, Methods of study. Journal of Educational Psychology, 1935, 26, 30-40.

Paden, D. W., & Moyer, M. E. The teaching of the first course in Economics. In A. L. Welsh (Ed.), Research papers in economic education. New York: Joint Council on Economic Education, 1972.

Parsons, J. A. Personalized system of instruction: Theory, research, and practice. Urbana: University of Illinois, mimeo.

Smock, H. R., & Myrow, D. L. Can "freedom to learn" be instituted? A preliminary survey of the undergraduate open seminar 199. Urbana: Office of Instructional Resources Report 121, University of Illinois, 1971.

Washburne, C. W. Educational measurements as a key to individualizing instruction and promotions. Journal of Educational Research, 1922, 5, 195-206.

Witters, D. R., & Kent, G. W. Teaching without lecturing: Evidence in the case for individualized instruction. The Psychological Record, 1972, 22, 169-175.

DISTRIBUTION LIST

Chief of Naval Operations (OP-987E)
Chief of Naval Personnel (Pers-10c)
Chief of Naval Education and Training (CNET N-33)
Chief of Naval Technical Training
Chief of Naval Material (NMAT-030B)
Chief of Naval Research (Code 458) (2)
(Code 450) (4)
Commanding Officer, Naval Training Equipment Center
Commanding Officer, Navy Medical Neuropsychiatric Research Unit
Commanding Officer, Naval Aerospace Medical Institute
Commanding Officer, Naval Education and Training Program Development Center
Commanding Officer, Naval Development and Training Center (Code 0120)
Superintendent, Naval Academy
Superintendent, Naval Postgraduate School
Superintendent, U. S. Military Academy
Superintendent, U. S. Air Force Academy
Superintendent, U. S. Coast Guard Academy
Army Research Institute for Behavioral and Social Sciences
Personnel Research Division, Air Force Human Resources Laboratory (AFSC), Lackland Air Force Base
Occupational Research Division, Air Force Human Resources Laboratory (AFSC), Lackland Air Force Base
Technical Training Division, Air Force Human Resources Laboratory (AFSC), Lowry Air Force Base
National Research Council
National Science Foundation
Science and Technology Division, Library of Congress
Defense Documentation Center (12)